TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Singh, S., Kashiwagi, R., Tanji, K., Roga, W., Bhatti, D., Takeoka, M., & Elkouss, D. (2026). Fault-tolerant modular quantum computing with surface codes using single-shot emission-based hardware. *Physical Review Applied*, *26*(2), Article 024005. https://doi.org/10.1103/6mqy-sd3d

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Fault-tolerant modular quantum computing with surface codes using single-shot emission-based hardware

Siddhant Singh,[1,*] Rikiya Kashiwagi,[2,3] Kazufumi Tanji,[3] Wojciech Roga,[3] Daniel Bhatti,[4] Masahiro Takeoka,[3,5,†] and David Elkouss[1,4,‡]

[1]*QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*
[2]*Graduate School of Informatics, Kyoto University, 36-1 Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
[3]*Department of Electronics and Electrical Engineering, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*
[4]*Networked Quantum Devices Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan*
[5]*National Institute of Information and Communications Technology (NICT), Koganei, Tokyo 184-8795, Japan*

(Received 21 January 2026; revised 3 April 2026; accepted 1 July 2026; published 5 August 2026)

Fault-tolerant modular quantum computing requires stabilizer measurements across the modules in a quantum network. For this, entangled states of high quality and rate must be distributed. Currently, two main types of entanglement distribution protocols exist, namely emission-based and scattering-based, each with its own advantages and drawbacks. On the one hand, scattering-based protocols with cavities or waveguides are fast but demand stringent hardware such as high-efficiency integrated circulators or strong waveguide coupling. On the other hand, emission-based platforms are experimentally feasible but have so far relied on Bell-pair fusion with extensive use of slow two-qubit memory gates, limiting thresholds to approximately 0.16%. Here, we consider a fully distributed surface code using emission-based entanglement schemes that generate GHZ states in a single shot, i.e., without the need for Bell-pair fusions. We show that our optical setup produces Bell pairs, W states, and GHZ states, enabling both memory-based and optical protocols for distilling high-fidelity GHZ states with significantly improved success rates. Furthermore, we introduce protocols that completely eliminate the need for memory-based two-qubit gates, achieving thresholds of approximately 0.19% with modest hardware enhancements, increasing to above approximately 0.24% with photon-number-resolving detectors. These results show the feasibility of emission-based architectures for scalable fault-tolerant operation.

DOI: 10.1103/6mqy-sd3d

I. INTRODUCTION

Scaling quantum computation to execute useful algorithms requires millions of physical qubits [1,2]. The modular quantum computing paradigm, comprising independent yet interconnected quantum modules, addresses this scalability challenge [3–10]. To enable this modularity in practice, several methods have been proposed to allow qubits in separate modules to interact with high success rates and fidelity [11–17]. These architectures must incorporate quantum error correction (QEC) to suppress errors [18–29]. Because of its high error-correction threshold of close to half a percent, the surface code has emerged as a leading candidate [30–35], guiding both algorithm development and hardware co-design [36–40]. Recent literature has extended these QEC techniques to modular systems, proposing a diverse range of implementations [28,41–50].

In this context, various entanglement generation schemes have been introduced and partially demonstrated [6,51–64] for color centers [65,66], neutral atoms [67–69], and trapped ions [70,71]. One such scheme, already demonstrated experimentally, is the emission-based approach [72,73]. It generates entanglement between different modules by measuring photons emitted from each module's communication (emitter) qubit and transmitted through an optical fiber. Then one builds a modular QEC architecture by using the module's memory qubits as code data qubits.

*Contact author: siddhant.singh@tudelft.nl
†Contact author: takeoka@elec.keio.ac.jp
‡Contact author: david.elkouss@oist.jp

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Typically, these memory qubits are also controlled via the emitter qubit.

A fully distributed surface-code experiment requires degree-four nearest-neighbor connectivity to measure stabilizers [6,41,43]. One approach is to distill and fuse Bell pairs to create high-fidelity four-partite GHZ states as a resource [6,41,43]. However, the two-qubit gates required for fusion and distillation are a bottleneck in modular architectures because emitter-qubit gates can be orders of magnitude faster than memory-qubit gates, as is the case for color centers [65,66]. Although memory-qubit gate speed can also be improved through optimization [74] or alternative gate protocols [75], enhancing gate speed while maintaining high fidelity is more difficult. Consequently, the performance of these fusion-based schemes was estimated to saturate around a gate-error threshold of 0.16% even with hardware improvements [6]. As an alternative, several scattering-based schemes have been proposed to avoid this limitation, but they demand highly complex hardware that is unlikely to be available in the near term [6].

To overcome these bottlenecks, here we consider a recent generalization of the bipartite entanglement emission-based scheme to multiple parties [76–78]. Such extensions avoid Bell-pair fusion and enable the direct distribution of multipartite entanglement in a single shot. However, the performance of these direct schemes on realistic hardware remains unknown, as does the potential for more efficient protocols based on this single-shot setup that exploit the existing setup to produce high-fidelity GHZ states. We pose the following questions to investigate whether the single-shot scheme can outperform the fusion-based schemes. Which near-term upgrade yields the largest threshold gain? Do thresholds scale continuously with hardware, unlike the fusion-based schemes? And what concrete requirements enable early fault-tolerant demonstrations?

In this work, we put forward these single-shot protocols by considering a four-module setup designed to generate entangled states such as raw GHZ states, Bell pairs, and W states between modules in a single-shot run of the setup. We propose protocols that use these states to produce high-fidelity GHZ states either through limited distillation using quantum memory gates or through purely optical distillation without the use of memory gates. We benchmark the best protocols for modular error correction by identifying fault-tolerant regions for realistic hardware parameters of color centers in diamond. These parameters include losses in optical components, photon emission and detection quality, and circuit-level noise in qubit operations.

Our results show that the existing emission-based hardware can support modular error correction and clarify the hardware requirements for achieving the surface-code error thresholds competitive with monolithic architectures. Notably, our optical distillation protocols are fast and enable direct GHZ generation without fusion, yielding code thresholds of 0.19% for non-photon-resolving detectors and 0.24% for number-resolving detectors with only modest improvements in hardware parameters. These thresholds scale with improving hardware parameters, unlike those of the Bell-pair fusion-based schemes. This establishes a clear pathway toward scalable modular architectures with emission-based hardware. This article is organized as follows. We present the physical setup for modular QEC in Sec. II. Section III then describes the novel protocols in the noiseless regime that are based on this single-shot emission setup for high-fidelity GHZ generation. Finally, we present the protocols' hardware performance with full noise, fault-tolerance hardware requirements, and threshold results in Sec. IV.

II. PHYSICAL SETUP FOR MODULAR QEC WITH SINGLE-SHOT EMISSION-BASED HARDWARE

A. Architecture comparison and overview

Recent progress in modular fault-tolerant quantum computing has produced a diverse landscape of architectural proposals, which differ substantially in their resource assumptions and operational regimes. Lattice surgery and transversal gate methods [79] and constant-rate entanglement distillation schemes [80] adopt a semidistributed architecture. Each module hosts a significant portion of a logical qubit or an entire logical block, and nonlocal operations are performed at the logical level between modules. This substantially relaxes the demands on intermodule link fidelity and rate, at the cost of requiring many physical qubits, high-fidelity two-qubit gates, and large local memory resources per module. Similarly, the fault-tolerant optical interconnect scheme of Ref. [48] operates at the logical level on neutral-atom arrays with Rydberg gates, targeting the high-local-resource regime.

The present work operates in a fundamentally different regime. We consider a fully distributed modular architecture in which each module hosts exactly one data qubit and a small number of distillation-ancilla memory qubits, with the surface code distributed at the single-qubit level and all stabilizer measurements requiring intermodule entanglement. Crucially, our formalism reduces modular fault tolerance to the reliable generation of high-fidelity multipartite entangled states for stabilizer measurements, which are the fundamental primitive underlying all higher-level operations, including lattice surgery and transversal gates. Any modular code architecture that requires nonlocal stabilizer measurements can therefore directly leverage the protocols introduced here. This minimal-resource-per-module regime directly targets color-center platforms: the fault-tolerant operation of a logical qubit within a single NV-center device has already been demonstrated [81], and the present work is the natural extension of that result to the modular setting. This addresses the intermodule entanglement generation bottleneck that must be overcome

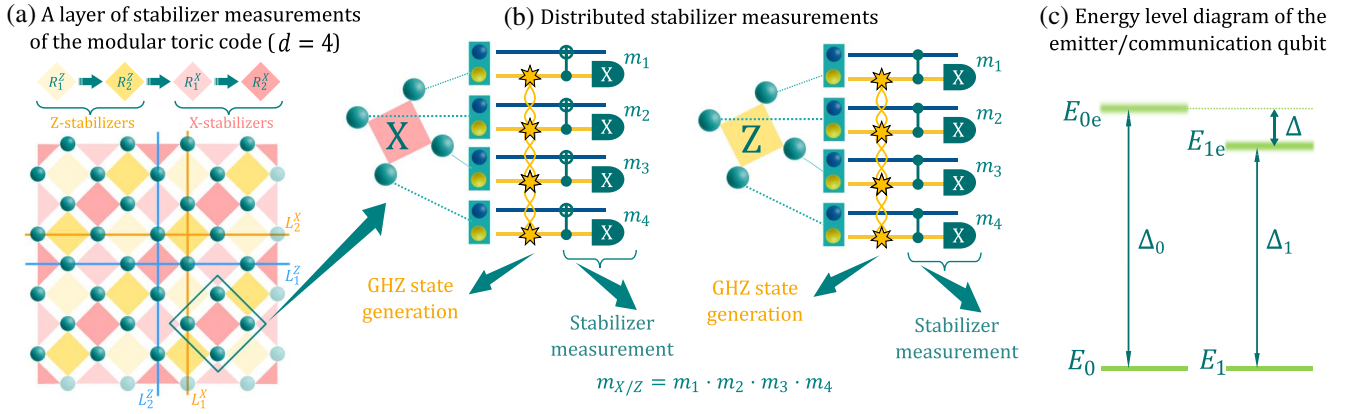


FIG. 1. Modular toric surface code. (a) A direct translation of the monolithic surface code of distance $d = 4$ into a modular architecture, where each module (teal circles) hosts a data qubit and a communication qubit (dark blue and yellow circles in (b), respectively). Each stabilizer spans four modules, which share GHZ states generated using the proposed single-shot emission-based protocols. As each module has only one communication qubit (yellow), simultaneous measurement of neighboring stabilizers is not possible. Thus, the stabilizers are measured in four sequential rounds $R^{X,Z}_{1,2}$, as shown. Logical operators are also shown for both encoded logical qubits. (b) Distributed stabilizer measurement circuits using a GHZ state in the form $|\Phi_4^+\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$. The stars (in yellow) connected via links denote entanglement generation between those communication qubits, either as a raw GHZ state or as a distilled GHZ state using distillation protocols. The joint parity is extracted by local operations within each module, followed by classical postprocessing. The measurement outcomes for the stabilizer syndrome are calculated as products of individual outcomes. (c) Energy level diagram of the communication (emitter) qubit, showing the ground states E_0 and E_1 (corresponding to dark and bright states) and the excited states E_{0e} and E_{1e}, along with their associated linewidths. For the noise model (see Appendix A), this is the emitter model followed in this work.

to scale beyond a single module. Our single-shot emission-based protocols are specifically designed for this regime, eliminating memory-gate requirements entirely in the optical distillation protocols, with hardware requirements fully compatible with demonstrated NV-center devices.

B. Single-shot modular architecture

We consider a distributed quantum architecture in which multiple modules are interconnected through optical links that enable entanglement generation on demand. For clarity, we describe the setup in two parts: (i) the quantum error correction machinery implementing the surface code in the memory, and (ii) the optical interface that establishes entanglement between modules to support these operations. These links are assumed to be short-distance links or rack-size assembled quantum computer-scale or integrated photonic waveguides. Each module contains a communication qubit, responsible for emitting photons, mediating intermodule entanglement, and acting as a control for local gates, together with several memory qubits. The memory qubits serve either as data qubits storing the logical information or as auxiliary qubits assisting in GHZ-state generation, depending on the protocol [6,41–43]. Only the communication qubit can be directly measured, while memory qubits are read out indirectly via SWAP operations. The physical platform we consider is inspired by color centers in diamond [66,82]. By arranging such modules into a two-dimensional code layout, one realizes a modular surface code capable of performing stabilizer measurements through distributed multiqubit entanglement (see Fig. 1). Our construction follows the weight-4 architecture of Ref. [6], exemplified by a modular toric code with periodic boundaries; planar variants without periodicity can be defined analogously. Stabilizer measurements are carried out using GHZ states distributed across modules via the optical links. Measuring the GHZ qubits yields correlated outcomes corresponding to parity checks, which are then processed by the Union-Find decoder to identify and apply the appropriate corrections.

For the optical interfaces, we consider a setup that allows GHZ generation for stabilizer measurements described above. We refer to this setup as a "single-shot" emission-based scheme. As we will show, this setup allows GHZ state generation between the communication qubits in one single run of the experiment. The focus is on the four-module unit cell (see Fig. 2) that is crucial for all stabilizer measurements. We first describe the hardware parameters and the relevant error sources in the emission-based setup. Subsequently, we present the different types of entanglement links that can be generated in the experiment. Each module is equipped with a local laser or microwave control, which can manipulate the communication qubit by applying local operations or by exciting it for emission. The initialization parameter for the communication-qubit state is the bright-state parameter α, which controls the population of the ground state $|0\rangle$ (with energy E_0) and the excited (bright) state $|1\rangle$ (with energy E_1) [the

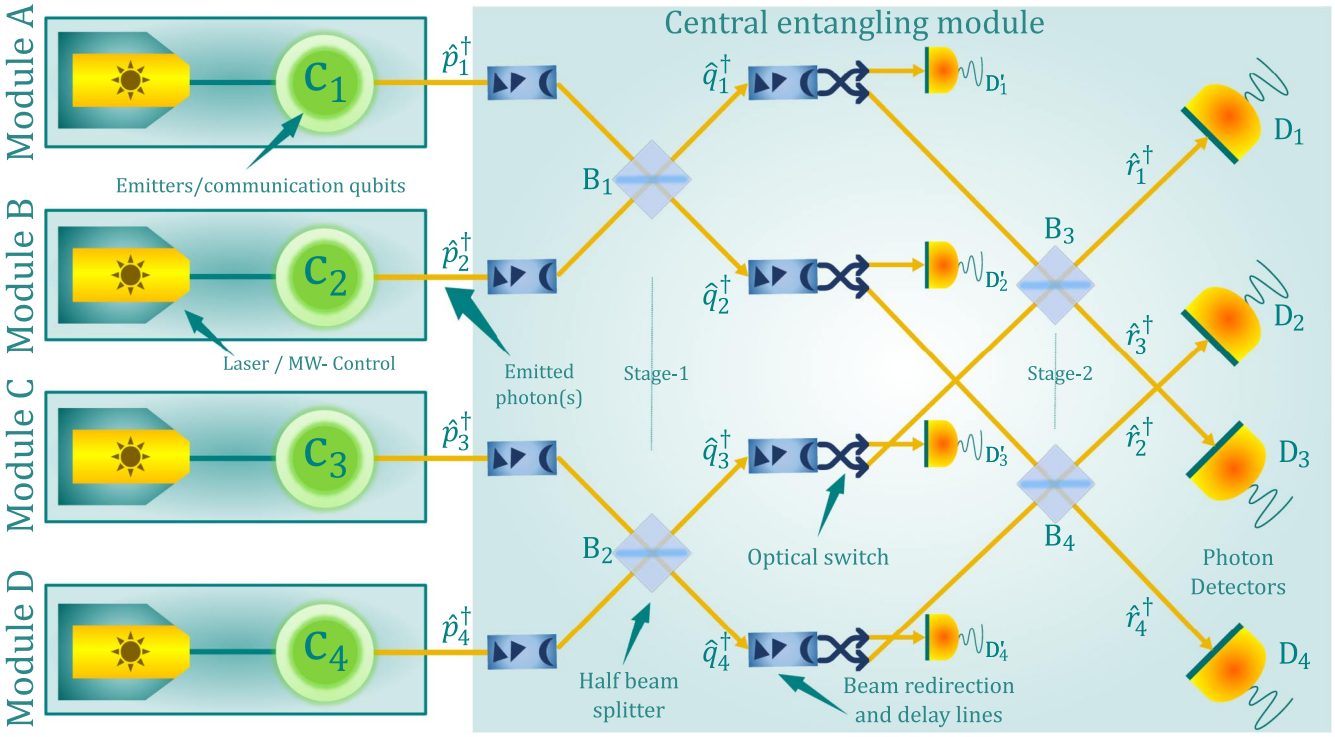


FIG. 2. Hardware setup for the emission-based scheme that generates entanglement links-Bell pairs, W states, and GHZ states of weight-4. Emitter (communication) qubits c_1, c_2, c_3, c_4 are excited with a laser pulse to emit photons. These photons propagate through a network of beam splitters (stage 1 and stage 2), which mix the modes of all emitted photons. We observe detection patterns as photon counts at detectors $\mathrm{D}_1, \mathrm{D}_2, \mathrm{D}_3, \mathrm{D}_4$ or at detectors $\mathrm{D}'_1, \mathrm{D}'_2, \mathrm{D}'_3, \mathrm{D}'_4$ depending on which states are created. Postselecting on specific patterns generates the target states. Creation operators for the photonic number states $\hat{p}_i^\dagger, \hat{q}_i^\dagger$ and $\hat{r}_i^\dagger$ $(i = 1, 2, 3, 4)$ are indicated at various input and output ports of the beam splitters.

energy level diagram of the emitter is shown in Fig. 1(c)]. The local communication-qubit state can be initialized with a fidelity F_{prep} and amplitude controlled by α. A subsequent local control pulse can then be used to emit a photon from the excited state, which becomes entangled with the communication qubit. Ideally, this process generates the emitter–photon entangled state $\sqrt{1-\alpha}\,|0\rangle|0_{\mathrm{ph}}\rangle + \sqrt{\alpha}\,|1\rangle|1_{\mathrm{ph}}\rangle$, with $|0_{\mathrm{ph}}\rangle$ ($|1_{\mathrm{ph}}\rangle$) being the zero-photon (single-photon) Fock state. The parameter α is tunable in the experiment.

During an excitation attempt, more than one photon may be emitted, resulting in a double-excitation error. If this unwanted photon is lost to the environment, it can induce dephasing in the state of the emitter. This effect was modeled in Ref. [6] by a dephasing probability p_{EE}. In this work, we neglect this parameter under the following assumptions: The emitter is placed in an optical resonator such as a cavity [83], the excitation pulse is shorter than the cavity-enhanced decay time, and the excitation pulse is optimized so that other transitions undergo a full cycle and return to the ground state [84,85].

As every photon propagates through an optical fiber, it may be lost due to transmission inefficiencies or may not be registered by an imperfect detector. We capture both effects using an effective photon detection efficiency, η_{ph}. Because the physical distance from the emitter modules to the central entangling module is much larger than the distance from the beam splitters to the detectors, we assume photon loss occurs before the beam splitters. Finally, photons emitted from different modules are not perfectly identical, so their interference at the beam splitter is not ideal. We model this effect with a two-photon visibility factor μ. Full details of these noisy hardware parameters, their modeling, and their effects on the quantum states are provided in Appendix A. We also take the detectors' number-resolving capability as a hardware parameter, and we model both photon-number-resolving (PNR) detectors and non-PNR detectors.

III. GHZ PROTOCOLS FOR SINGLE-SHOT EMISSION SCHEME

Based on the emission-based setup described in the previous section, we propose several entanglement-generation protocols for high-fidelity GHZ states. In this section, we present the noiseless working principles of these protocols, and in Appendix A we provide a recipe for constructing full noisy analytical simulations by explaining the noise sources and the error channels we used for their modeling. First, we show how the elementary

TABLE I. Protocol naming convention.

Short label	Full name	Type
SC-Bell	Single-click Bell link	Elementary
DC-Bell	Double-click Bell link	Elementary
Raw-W	Raw W link	Elementary
Raw-GHZ	Raw GHZ link	Elementary
MD-BG(SC)	MD: SC-Bell + GHZ	Memory distillation
MD-BG(DC)	MD: DC-Bell + GHZ	Memory distillation
MD-WG	MD: W + GHZ	Memory distillation
MD-WW	MD: $W + W$	Memory distillation
MD-GG	MD: GHZ + GHZ	Memory distillation
OD-GHZ	OD: GHZ + GHZ	Optical distillation
OD-W	OD: $W + W$	Optical distillation

entangled states can be generated using this setup, and then we describe the high-fidelity GHZ protocols in detail and calculate their success probabilities and the fidelities of the resulting GHZ states. We classify the protocols into two categories: (i) Memory-based distillation (MD) protocols that apply two-qubit gates on the memory qubits of the modules to perform distillation. (ii) Optical distillation (OD) protocols that avoid any two-qubit gates on the memory qubits and instead rely on multiple rounds of emission operations and local qubit rotations on the communication qubits.

Throughout this paper, we adopt a unified naming convention for all protocols. The Bell, W, or GHZ elementary links are referred to by their names. Memory distillation protocols are labeled MD-XY, where XY encodes the two input resource states. The first state is the distilling state, and the second is the distilled state. Optical distillation protocols are labeled OD-XY, where XY denotes the states used in the first and second rounds, respectively. Since all OD and MD protocols produce a GHZ output, we do not explicitly state this in the protocol naming. The full protocol index is given in Table I.

A. Creating elementary states

1. Creating a Bell state

As shown in Fig. 2, one can use detectors D_1', D_2' (D_3', D_4') to generate entangled Bell-pair links between the subsystems A and B (C and D). The optical switches reroute the incoming photons to the selected set of detectors. We consider both the single-click and double-click protocols for generating Bell-pair links between these modules. Both protocols were studied in Ref. [6], and here we restate only their noiseless working principles.

First, each module A and B (C and D) generates a local emitter-photon state. The total state is

$$|\psi_{\text{init}}\rangle = (\sqrt{1-\alpha}\,|00_{\text{ph}}\rangle + \sqrt{\alpha}\,|11_{\text{ph}}\rangle)^{\otimes 2}, \tag{1}$$

where the subscript 'ph' labels the number state of the photon(s). Going through a balanced two-mode beam splitter, the transformation for the photonic creation operators can be modeled as

$$\begin{pmatrix} \hat{p}_1^\dagger \\ \hat{p}_2^\dagger \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \hat{q}_1^\dagger \\ \hat{q}_2^\dagger \end{pmatrix}, \tag{2}$$

where $\hat{p}_i^\dagger$ and $\hat{q}_j^\dagger$ are photonic creation operators for input and output ports for stage-1 beam splitters (see Fig. 2). This transformation maps the photon number states from input ports to output ports as follows:

$$|0_{\text{ph}}0_{\text{ph}}\rangle_{\text{in}} \to |0_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}}, \tag{3}$$

$$|0_{\text{ph}}1_{\text{ph}}\rangle_{\text{in}} \to \frac{1}{\sqrt{2}}(|1_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}} - |0_{\text{ph}}1_{\text{ph}}\rangle_{\text{out}}), \tag{4}$$

$$|1_{\text{ph}}0_{\text{ph}}\rangle_{\text{in}} \to \frac{1}{\sqrt{2}}(|0_{\text{ph}}1_{\text{ph}}\rangle_{\text{out}} + |1_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}}), \tag{5}$$

$$|1_{\text{ph}}1_{\text{ph}}\rangle_{\text{in}} \to \frac{1}{\sqrt{2}}(|2_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}} - |0_{\text{ph}}2_{\text{ph}}\rangle_{\text{out}}). \tag{6}$$

Therefore, the joint emitter-photon state of the two modules becomes

$$\begin{aligned} |\psi_{\text{final}}\rangle = {} & (1-\alpha)|00\rangle|0_{\text{ph}}0_{\text{ph}}\rangle \\ & + \sqrt{\frac{\alpha(1-\alpha)}{2}}(|01\rangle + |10\rangle)|0_{\text{ph}}1_{\text{ph}}\rangle \\ & - \sqrt{\frac{\alpha(1-\alpha)}{2}}(|01\rangle - |10\rangle)|1_{\text{ph}}0_{\text{ph}}\rangle \\ & + \frac{\alpha}{\sqrt{2}}|11\rangle(|2_{\text{ph}}0_{\text{ph}}\rangle - |0_{\text{ph}}2_{\text{ph}}\rangle). \end{aligned} \tag{7}$$

Observe that if we now obtain a single-photon click in detector D_1' (D_2'), we obtain the perfect Bell state of the emitters as $|\psi^{\text{PNR}}_{\text{emitters,Bell}}\rangle = |\Psi^{\mp}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \mp |10\rangle)$. For any other detection pattern, we abort the protocol and repeat until success (RUS). As pointed out above, the detectors may not be PNR, and two photons can be registered as a false single click, which adds the term $|11\rangle$ to the emitters' density matrix. In the case of non-PNR detectors, for instance, when D_2' clicks, we obtain a mixed state (reduced density matrix state after tracing out the photons)

$$\rho^{\text{non-PNR}}_{\text{emitters,Bell}} = (1-\alpha)|\Psi^+\rangle\langle\Psi^+| + \alpha|11\rangle\langle 11|. \tag{8}$$

Note that here we have neglected dark counts, which correspond to detectors registering a false click. Since we assume sufficiently fast remote entanglement generation, the probability of dark counts is very small and can be safely neglected [86]. Looking at Eq. (8), it is clear that in the non-PNR case, one needs a vanishingly small α to achieve a high-fidelity Bell pair.

To further improve the Bell-pair fidelity with non-PNR detectors, we employ the double-click (DC) protocol (also known as Barrett-Kok protocol [73]). Local bit-flip operations are applied to the emitters, flipping the undesired two-photon-emitting term ($|11\rangle\langle 11|$) into the nonemitting term ($|00\rangle\langle 00|$), leaving the term $|\Psi^+\rangle\langle\Psi^+|$ unchanged. Another round of emission is then executed, yielding a perfect Bell pair upon another single click in one of the detectors, even with non-PNR detectors. Note that this requires two successive rounds of detection patterns, implemented in an RUS manner. And the overall success rate will depend quadratically on the probability of such single clicks. Full noisy analytical results for Bell-pair creation were previously reported in Ref. [6], and we follow this model identically with the same noise model.

2. Creating a W state

Now we focus on how to craft weight-4 W states using the same experimental setup. In this case, the optical switches reroute the incoming photons to the stage-2 beam splitters, after which they are detected at the final detectors D_1, D_2, D_3, D_4 (see Fig. 2). Local emitter-photon states are created within each module A, B, C, and D, analogous to Eq. (1). The full unitary operation of all the beam splitters, mapping the input creation operators $\hat{p}_i^\dagger$ to the output operators $\hat{r}_j^\dagger$, can be modeled as (see Appendix B) [77]

$$\begin{pmatrix} \hat{p}_1^\dagger \\ \hat{p}_2^\dagger \\ \hat{p}_3^\dagger \\ \hat{p}_4^\dagger \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} \hat{r}_1^\dagger \\ \hat{r}_2^\dagger \\ \hat{r}_3^\dagger \\ \hat{r}_4^\dagger \end{pmatrix}. \tag{9}$$

The emitter-photon state for all modules is

$$|\psi_{\text{init}}\rangle = (\sqrt{1-\alpha}\,|00_{\text{ph}}\rangle + \sqrt{\alpha}\,|11_{\text{ph}}\rangle)^{\otimes 4}. \tag{10}$$

The photonic terms from this input state can be written in the shorthand notation as

$$(\hat{p}_1^\dagger)^{p_1}(\hat{p}_2^\dagger)^{p_2}(\hat{p}_3^\dagger)^{p_3}(\hat{p}_4^\dagger)^{p_4}|0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}\rangle_{\text{in}}, \tag{11}$$

where $p_1, p_2, p_3, p_4 \in \{0, 1\}$ are integers that define the number of photons in each input mode. We assume that at most one photon is emitted per module and contributes to the output detection pattern. Any additional photon from a double-excitation is either lost due to dephasing or occurs with vanishing probability, as justified by the physical model. Therefore, double-excitation errors can be safely ignored. Inducing the transformation of the beam splitter, we find the photonic output state

$$\begin{aligned}&(\hat{p}_1^\dagger)^{p_1}(\hat{p}_2^\dagger)^{p_2}(\hat{p}_3^\dagger)^{p_3}(\hat{p}_4^\dagger)^{p_4}|0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}\rangle_{\text{in}} \\ &\quad\to \frac{1}{2^{p_1+p_2+p_3+p_4}}(\hat{r}_1^\dagger + \hat{r}_2^\dagger + \hat{r}_3^\dagger + \hat{r}_4^\dagger)^{p_1} \\ &\qquad\times (\hat{r}_1^\dagger - \hat{r}_2^\dagger + \hat{r}_3^\dagger - \hat{r}_4^\dagger)^{p_2}(\hat{r}_1^\dagger + \hat{r}_2^\dagger - \hat{r}_3^\dagger - \hat{r}_4^\dagger)^{p_3} \\ &\qquad\times (\hat{r}_1^\dagger - \hat{r}_2^\dagger - \hat{r}_3^\dagger + \hat{r}_4^\dagger)^{p_4}|0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}},\end{aligned} \tag{12}$$

where we have ignored the prefactors depending on α and only shown the photonic degrees of freedom [see Eq. (11)]. In general, many detection patterns are possible, corresponding to various input combinations. However, to create a W state, we only postselect on the detection pattern where we get a single click in one of the detectors. Note that these detection patterns and success rates are actually modeled via the positive operator-valued measure (POVM)-based Kraus operators described in Appendix B, which takes into account photon indistinguishability factors as described in Appendix A 2 d.

Say that we get a click from a single photon at detector D_1, meaning that we get the output state $\hat{r}_1^\dagger|0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}0_{\text{ph}}\rangle_{\text{out}}$. This pattern is possible via a superposition of different single excitations from various emitter states [using Eq. (12)], creating the entangled state of the emitters

$$\begin{aligned}&|\psi^{\text{PNR}}_{\text{emitters,W}}\rangle \\ &\quad= \frac{1}{2}(|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle) = |W_4\rangle.\end{aligned} \tag{13}$$

When the detectors are perfectly PNR, a single click in D_1 (or any single click in one of the detectors with appropriate postdetection local rotations) yields the desired state. With non-PNR detectors, unwanted higher-photon terms also contribute to the final state. For example, the detector D_1 cannot distinguish between the photon-number terms $r_1^\dagger$, $(r_1^\dagger)^2$, $(r_1^\dagger)^3$, and $(r_1^\dagger)^4$. This introduces noise, and we obtain the following state:

$$\begin{aligned}&\rho^{\text{non-PNR}}_{\text{emitters,W}} \\ &= \frac{1}{\mathcal{N}_W}\left(\alpha(1-\alpha)^3|W_4\rangle\langle W_4| + \frac{3\alpha^2(1-\alpha)^2}{4}|E_2\rangle\langle E_2|\right. \\ &\quad\left. + \frac{3\alpha^3(1-\alpha)}{8}|W_4'\rangle\langle W_4'| + \frac{3\alpha^4}{32}|1111\rangle\langle 1111|\right),\end{aligned} \tag{14}$$

where the first term is the desired W state, resulting from a true single-photon detection in D_1 and $\mathcal{N}_W$ denotes a normalization factor. The second term comes from a single click in D_1 triggered by two unresolved photons, with $|E_2\rangle = \frac{1}{\sqrt{6}}(|0011\rangle + |1100\rangle + |0101\rangle + |1010\rangle + |0110\rangle + |1001\rangle)$, the third term comes from three unresolved photons in D_1, with $|W_4'\rangle = X_1X_2X_3X_4|W_4\rangle$, and the fourth

term comes from four unresolved photons in D_1. In this case, the fidelity of the desired W state is relatively lower and dependent upon the bright-state parameter. We get the highest fidelity with the lowest value of α.

3. Creating a GHZ state

Creating a GHZ state works analogously to the creation of the W state. The difference is in the detection pattern that we postselect on. To create a GHZ state, we demand exactly two single-photon detections in different detectors, e.g., D_1 and D_2. Based on Eq. (12), we find two two-photon input terms that can lead to this detection pattern, i.e., $(p_1, p_2, p_3, p_4) = (1,0,1,0), (0,1,0,1)$. Explicitly, this results in the terms

$$\begin{aligned}\hat{p}_1^\dagger \hat{p}_3^\dagger |0_{\rm ph}0_{\rm ph}0_{\rm ph}0_{\rm ph}\rangle_{\rm in} &\to \frac{1}{4}(\hat{r}_1^{\dagger 2} + \hat{r}_2^{\dagger 2} - \hat{r}_3^{\dagger 2} - \hat{r}_4^{\dagger 2} \\ &+ 2\hat{r}_1^\dagger \hat{r}_2^\dagger - 2\hat{r}_3^\dagger \hat{r}_4^\dagger)|0_{\rm ph}0_{\rm ph}0_{\rm ph}0_{\rm ph}\rangle_{\rm out}, \\ \hat{p}_2^\dagger \hat{p}_4^\dagger |0_{\rm ph}0_{\rm ph}0_{\rm ph}0_{\rm ph}\rangle_{\rm in} &\to \frac{1}{4}(\hat{r}_1^{\dagger 2} + \hat{r}_2^{\dagger 2} - \hat{r}_3^{\dagger 2} - \hat{r}_4^{\dagger 2} \\ &- 2\hat{r}_1^\dagger \hat{r}_2^\dagger + 2\hat{r}_3^\dagger \hat{r}_4^\dagger)|0_{\rm ph}0_{\rm ph}0_{\rm ph}0_{\rm ph}\rangle_{\rm out}. \end{aligned} \tag{15}$$

Postselecting based on this detection pattern prepares the emitters in the perfect GHZ state

$$|\psi^{\rm PNR}_{\rm emitters,GHZ}\rangle = \frac{1}{\sqrt{2}}(|0101\rangle - |1010\rangle) = |\Psi_4^-\rangle, \tag{16}$$

which can be converted to the $|\Phi_4^+\rangle$GHZ state with local operations. We call this the "Raw GHZ" state. Note that this perfect GHZ generation is only possible in the noiseless case with PNR detectors.

When we have non-PNR detectors, we must also include contributions from the output terms corresponding to $\hat{r}_1^{\dagger 2}\hat{r}_2^\dagger$, $\hat{r}_1^{\dagger 3}\hat{r}_2^\dagger$, $\hat{r}_1^\dagger \hat{r}_2^{\dagger 2}$, $\hat{r}_1^\dagger \hat{r}_2^{\dagger 3}$, and $\hat{r}_1^{\dagger 2}\hat{r}_2^{\dagger 2}$ which correspond to detecting up to a total of four photons in both detectors combined. We collect the contribution due to all these terms on the emitters' state, which gives us the mixed state

$$\begin{aligned}&\rho^{\rm non\text{-}PNR}_{\rm emitters,GHZ} \\ &= \frac{1}{\mathcal{N}_{\rm GHZ}}\left(\frac{\alpha^2(1-\alpha)^2}{2}|\Psi_4^-\rangle\langle\Psi_4^-| + \frac{\alpha^3(1-\alpha)}{4}\rho_{1,2}\right. \\ &\left. + \frac{\alpha^4}{16}|1111\rangle\langle 1111|\right). \end{aligned} \tag{17}$$

where the fidelity is dependent on α and we get higher fidelity with lower α, up to some normalization factor $\mathcal{N}_{\rm GHZ}$. The second term corresponds to the noise term resulting from detecting a single-photon click and an unresolved two-photon click, with $\rho_{1,2}$ containing only terms with exactly three emitters in the excited state, and the third term results from measuring four photons unresolved at the two detectors. Three variations of the four-party GHZ state can be generated based on ${}^4C_2 = 6$ two-detector-click patterns. All of them can be transformed into a raw $|\Phi_4^+\rangle$GHZ state, which is needed (in our convention) for the modular stabilizer measurement.

B. Memory distillation protocols

Having established reliable methods for creating elementary entangled states, we now introduce a set of distillation protocols that build on these states. Reference [6] demonstrated that fusion-based architectures fail to achieve high thresholds because they rely on an excessive number of memory-based two-qubit gates. In contrast, the extent to which distillation can be effectively incorporated into an emission-based architecture has remained unclear. To address this gap, we investigate minimal distillation strategies for generating GHZ states from the elementary links described above. The proposed memory-distillation protocols are specifically adapted to the architectural quantum circuits of diamond color-center hardware. In each protocol, we first generate a base state in the form of an elementary link and swap it onto the auxiliary memory qubits within each module. We then generate an additional elementary link and employ a dedicated distillation circuit that, with a certain success probability, purifies the stored base state to yield a high-fidelity GHZ state.

1. MD: Bell + GHZ protocol

Having created a raw GHZ state, minimal-resource distillation can be achieved by employing Bell states. Bell states have higher success probabilities than directly creating a raw GHZ state. We generate two Bell states simultaneously in a RUS procedure between modules A–B and C–D to distill the raw GHZ state. Either the single-click or double-click protocol can be used for creating these Bell states, and we name the protocols accordingly. The raw GHZ state is first created and swapped to the auxiliary qubits, freeing the communication qubits. Simultaneous Bell-pair generation is then performed between modules A–B and C–D. Once both Bell pairs are successfully created, after potentially different waiting times, the distillation protocol is executed [see Fig. 3(a) for the quantum circuit]. First, CNOT gates are applied from the communication qubits to the auxiliary qubits, implementing a joint XX stabilizer measurement that projects the raw GHZ state. The communication qubits are then measured in the Z basis, with postselection on the required measurement outcomes and appropriate postmeasurement corrections. These desired measurement patterns are shown at the bottom of Fig. 3(a). Finally, the distilled GHZ state is swapped back to the communication qubits to proceed with the stabilizer measurement of the surface code.

We can understand how this protocol works by looking at the noiseless raw GHZ and Bell states [see Eqs. (8)

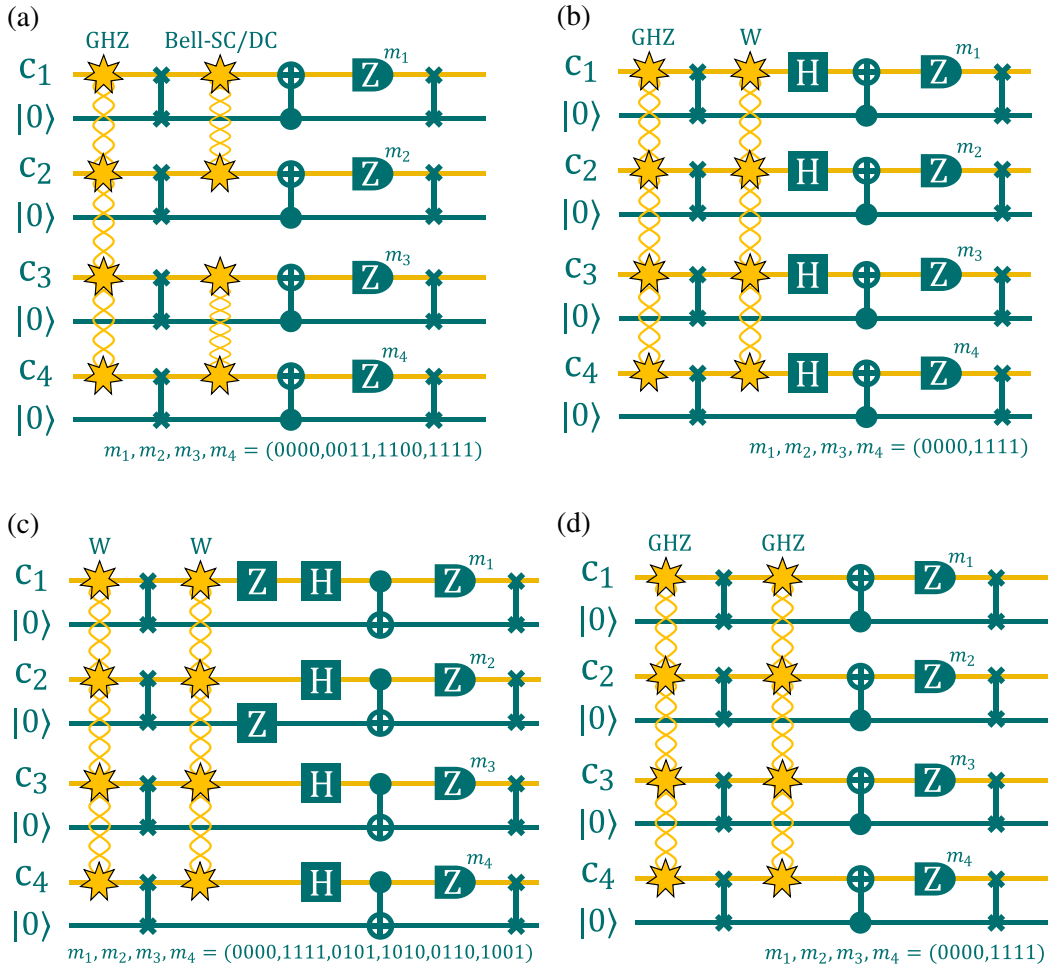


FIG. 3. Quantum circuits for memory-assisted distillation protocols. Communication qubits (yellow wires) act as photon-emitting qubits, which create entangled states, while auxiliary memory qubits (teal wires) assist with the storage of entangled states for the memory-assisted distillation. Elementary entanglement links are indicated by starred connections, labeled with the raw state produced (GHZ, Bell, or W). Postselection is performed on measurement outcomes m_1, m_2, m_3, m_4 with the indicated valid patterns, and the protocols are executed in an RUS manner as shown in the subfigures. (a) MD: SC/DC-Bell + GHZ: a base GHZ link is distilled with auxiliary Bell pairs. (b) MD: W + GHZ: a base GHZ link is distilled with auxiliary W state. (c) MD: $W + W$: two raw W states are locally rotated and distillation projects the base state into a GHZ state. (d) MD: GHZ + GHZ : two raw GHZ states are created, and the new one is used to distill the prior one.

and (17)]. Postselecting for the 1111 outcome after applying the CNOT gates can only result from having generated a perfect GHZ state and two perfect Bell states, thereby distilling a perfect GHZ state. On top of that, postselecting for 0000, 1100, and 0011, which are valid measurement patterns, also increases the success rate of the GHZ by yielding different GHZ states that can be converted to $|\Phi_4^+\rangle$ with appropriate local operations. Note that if the value of α is large, with non-PNR detectors as the only noise source, then post-selecting on the 1111 measurement pattern gives the highest fidelity, as it eliminates the higher excitation terms via distillation. However, in our simulations (see Sec. IV), which include multiple sources of noise with a very low value of α (for the highest fidelity of the elementary states), we choose other measurement patterns to increase the protocol success rate.

2. MD: W + GHZ protocol

In this protocol, we use a W state to distill a raw GHZ state. The raw state is first created using the RUS procedure and is then swapped onto the auxiliary memory qubits, freeing the communication qubits for another round of elementary link generation. A W state is subsequently created on the communication qubits, again using RUS. We perform a basis transformation on the W-state qubits using Hadamard gates

$$\begin{aligned}|\tilde{W}_4\rangle &= H_1H_2H_3H_4|W_4\rangle \\ &= \frac{1}{2}|W_4\rangle - \frac{1}{2}|W_4'\rangle + \frac{1}{\sqrt{2}}|\Phi_4^-\rangle. \end{aligned} \tag{18}$$

One can directly see that the rotated W state contains the GHZ state $|\Phi_4^-\rangle$ and can therefore be used for GHZ state

distillation. Again, the distillation protocol consists of applying CNOT gates from the auxiliary qubits to the communication qubits, followed by final measurements on the communication qubits. This sequence completes the protocol when the desired measurement pattern is obtained. Since both the rotated W state and the raw GHZ state [see Eq. (17)] contain terms with odd numbers of excitations, it becomes clear that this protocol cannot yield perfect GHZ states. Its advantage lies in the higher success rate for generating W states instead of GHZ states, which allows for lower values of α and, thereby, an increase in fidelity. The schematic of the protocol is depicted in Fig. 3(b).

3. MD: W + W protocol

This memory-based distillation protocol distills a GHZ state from solely using two W states. Similar to the above protocols, we first generate two raw W states (RUS). Then the states are locally rotated with Z and Hadamard rotations [as shown in Fig. 3(c)]. Assuming perfect W states, the rotated states take the following forms:

$$H_1H_2H_3H_4Z_1|W_4\rangle = X_1|\tilde{W}_4\rangle, \tag{19}$$

$$H_1H_2H_3H_4Z_2|W_4\rangle = X_2|\tilde{W}_4\rangle. \tag{20}$$

From Eq. (18), one can understand that both rotated W states contain all possible terms with even numbers of excitations coming from $X_i(|W_4\rangle - |W_4'\rangle)$. At the same time, they contain nonequivalent terms with odd numbers of excitations coming from $X_i|\Phi_4^-\rangle$. Applying CNOT gates and postselecting for six different patterns with even numbers of 0 s and 1 s yields all terms with even numbers of excitations, which are identical to a rotated GHZ state. Finally, rotating the distilled GHZ state transforms it into $|\Phi_4^+\rangle$. The schematic of the full protocol is depicted in Fig. 3(c).

4. MD: GHZ + GHZ protocol

The most intuitive distillation protocol consists of generating two raw GHZ states successively and distilling one from the other [78]. This can be understood from looking at the form of the noiseless raw GHZ state without PNR [see Eq. (17)]. Applying CNOT gates between two copies of this state and postselecting for 0000 or 1111 outcomes directly increases the weight of the GHZ state terms. Postselecting only for the 1111 outcome would even generate a perfect GHZ state in the noiseless scenario. We show the quantum circuit in Fig. 3(d) along with the measurement outcome patterns. Similar to the case of the Bell state to GHZ distillation protocol, the 1111 measurement pattern gives the highest fidelity when α is large, but due to the regime of our simulation parameters, we also include the 0000 measurement pattern.

C. Optical distillation protocols

We have proposed protocols that require minimal memory distillation in the previous section, meaning that each protocol involves only a single round of distillation. When gates are very noisy or slow, it becomes necessary to explore alternative approaches that can produce high-fidelity GHZ states without relying on memory-based two-qubit gates. Optical distillation protocols, such as the DC protocol, enable this by performing two successive rounds of photon emission and detection to filter out the desired state. The downside is that the overall success rate becomes the square of the individual success rates, which is disadvantageous in regimes with high loss or low effective photon detection efficiency. Nevertheless, we show that these protocols yield high-fidelity GHZ states and offer promising performance for modest values of photon detection efficiency.

1. OD: GHZ + GHZ protocol

Analogous to the DC version of the Bell-pair generation [73], we can generalize the single-shot raw-GHZ state generation protocol to a DC setup. Here, DC refers to the double-click protocol, which consists of two rounds of optical detection clicks with respective desired patterns. For this, we start by generating a raw GHZ state as discussed in Sec. III A 3. From Eq. (17), we observe that the raw GHZ state only contains noise terms with three or four excited communication qubits. Applying local bit-flip operations to all communication qubits, i.e., $X_1X_2X_3X_4$, changes the raw GHZ state to

$$\begin{aligned}\rho_{\text{emitters,GHZ}}^{\text{non-PNR}} \xrightarrow{X_1X_2X_3X_4} & \\ \times \frac{1}{\mathcal{N}_{\text{GHZ}}}\Bigg(&\frac{\alpha^2(1-\alpha)^2}{2}|\Psi_4^-\rangle\langle\Psi_4^-| + \frac{\alpha^3(1-\alpha)}{4}\rho_{1,2}' \\ &+ \frac{\alpha^4}{16}|0000\rangle\langle 0000|\Bigg).\end{aligned} \tag{21}$$

From this state, one can see that the GHZ terms remain unchanged. However, the undesired noise terms, i.e., $\rho_{1,2}$ and $|1111\rangle\langle 1111|$, are flipped to $\rho_{1,2}'$ and $|0000\rangle\langle 0000|$, which contain only one and zero excited qubits, respectively, and cannot emit more than a single photon. Executing a second round of emission and postselecting for the same two-detector patterns as in the raw state generation round generates a perfect GHZ state even without using PNR detectors. Using PNR detectors increases the theoretical success probability by a factor of 2 since one can also postselect for both photons arriving at the same detector. The protocol is summarized in Fig. 4.

2. OD: W + W protocol

Remarkably, one can also find a DC protocol generating GHZ states when starting from a raw W state. As in the

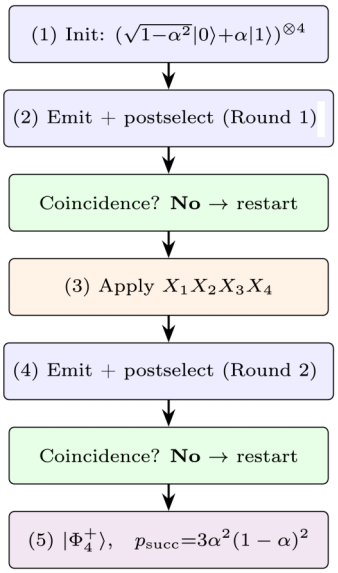


FIG. 4. OD-GHZ protocol operational summary (RUS). (1) All four communication qubits are initialized and excited to emit photons into the beam-splitter network. (2) A two-detector coincidence postselection heralds the raw-GHZ state [Eq. (17)]. (3) Local $X_1X_2X_3X_4$ gates map the noise terms into the low-excitation subspace [Eq. (21)]. (4) A second emission round is executed on the same device without reconfiguration. (5) Postselection on the same coincidence pattern heralds the purified GHZ state $|\Phi_4^+\rangle$ with $p_{\text{succ}} = 3\alpha^2(1-\alpha)^2$. A failed coincidence at either round triggers a full restart.

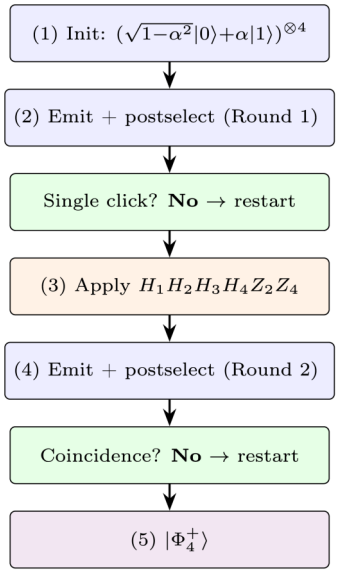


FIG. 5. OD-W protocol operational summary (RUS). (1) Initialize and emit photons into the beam-splitter network. (2) Single-detector click heralds the raw W state. (3) Local $H_1H_2H_3H_4Z_2Z_4$ rotations expose the GHZ component. (4) Second emission round with two-detector coincidence postselection. (5) Output approximates $|\Phi_4^+\rangle$; unlike OD-GHZ, a perfect GHZ state requires $\alpha \to 0$ as W-state noise terms cannot be fully suppressed. A failed postselection at either round triggers a full restart.

memory distillation protocols using W states, the reason for this lies in the rotated W state. Here, we are interested in a rotated W state with slightly different phases, i.e.,

$$H_1H_2H_3H_4Z_2Z_4|W_4\rangle = X_2X_4|\tilde{W}_4\rangle = \frac{1}{2}X_2X_4(|W_4\rangle - |W_4'\rangle) + \frac{1}{\sqrt{2}}|\Psi_4^-\rangle. \quad (22)$$

From this state, we can see that executing another round of emission and postselecting for two-detector (two-photon) coincidences using non-PNR (PNR) detectors generates a GHZ state. However, looking at Eq. (14), it becomes obvious that this protocol cannot generate a perfect GHZ state under realistic conditions. In contrast to the OD-GHZ protocol, the OD-W protocol cannot suppress the noise terms present in the raw W state completely. Since the Hadamard rotations of the noise terms also contain two excitation GHZ terms, they cannot be filtered out. Only in the case of a perfect raw W state ($\alpha \to 0$) can this protocol generate a close to perfect GHZ state. The operational summary of this protocol is presented in FIG. 5.

D. Summary of all protocols under noiseless hardware

We present the output results of all protocols in Table II. The success rates and fidelities of the output GHZ states are reported as functions of the tunable bright-state parameter α, for both PNR and non-PNR detector types. For elementary link generation and optical distillation protocols, we provide analytical expressions. For the expressions of success rates, we have taken into account all detection patterns of each round that contribute to the creation of the desired state, since all these output states can be locally rotated to yield the $|\Phi_4^+\rangle$GHZ state. For memory distillation protocols, the overall success rate depends on multiple independent RUS events for each link generation. We define the total protocol success rate as the ratio of successful protocol executions (n_{distill}) to the total number of simulation shots, n_{shots}. For each required elementary "state," the number of attempts until success is denoted by $n_{\text{attempts}}^{\text{state}}$. For each successful attempt, we compute the fidelity $F_{i,\text{GHZ}}$ of the output state and report the final protocol fidelity as the numerical average over all successful attempts. In our simulations, we use $n_{\text{shots}} = 10^4$, yielding statistical errors below 0.1% per data point.

IV. RESULTS

In this section, we assess whether single-shot entanglement generation protocols can outperform fusion-based schemes. To this end, we analyze all proposed single-shot emission-based protocols under full hardware noise and

TABLE II. Output state success rates and fidelity for various entanglement generation protocols with noiseless hardware. *Abbreviations*: SC, single-click; DC, double-click; PNR, photon-number resolving; $n_s = n_{\text{shots}}$—the total shots of the distillation protocol; $n_d = n_{\text{distil}}$—shots where the respective distillation pattern succeeds; $n_a = n_{\text{attempts}}$—to generate a respective elementary state via RUS (superscripts R, B, W, G denote Raw, Bell, W, GHZ, respectively); $F_{i,G} = F_{i,\text{distilled-GHZ}}$.

Protocol (detector type)	Success Rate	Fidelity
SC-Bell (PNR) and SC-Bell (non-PNR)	$2\alpha(1-\alpha)$ & $\alpha(2-\alpha)$	1 & $\sqrt{\frac{2(1-\alpha)}{2-\alpha}}$
DC-Bell (both)	$2\alpha(1-\alpha)$	1
W (PNR) and W (non-PNR)	$4\alpha(1-\alpha)^3$ & $\frac{\alpha(32-72\alpha+60\alpha^2-17\alpha^3)}{8}$	1 & $4\sqrt{\frac{2(1-\alpha)^3}{32-72\alpha+60\alpha^2-17\alpha^3}}$
GHZ (PNR) and GHZ (non-PNR)	$3\alpha^2(1-\alpha)^2$ & $\frac{3\alpha^2(5\alpha^2-12\alpha+8)}{8}$	1 & $\frac{2\sqrt{2}(1-\alpha)}{\sqrt{5\alpha^2-12\alpha+8}}$
MD: SC/DC-Bell + GHZ (both)	$(\frac{n_d}{n_s})/\sum_{i=1}^{n_s}[n_a^G + \max(n_a^{B1}, n_a^{B2})]_i$	$\frac{1}{n_s}\sum_{i=1}^{n_s} F_{i,G}$
MD: W + GHZ (both)	$(\frac{n_d}{n_s})/\sum_{i=1}^{n_s}[n_a^G + n_a^W]_i$	$\frac{1}{n_s}\sum_{i=1}^{n_s} F_{i,G}$
MD: $W + W$ (both)	$(\frac{n_d}{n_s})/\sum_{i=1}^{n_s}[n_a^{W1} + n_a^{W2}]_i$	$\frac{1}{n_s}\sum_{i=1}^{n_s} F_{i,G}$
MD: GHZ + GHZ (both)	$(\frac{n_d}{n_s})/\sum_{i=1}^{n_s}[n_a^{G1} + n_a^{G2}]_i$	$\frac{1}{n_s}\sum_{i=1}^{n_s} F_{i,G}$
OD: GHZ + GHZ (both)	$3\alpha^2(1-\alpha)^2$	1
OD: $W + W$ (PNR) and (non-PNR)	$2\alpha(1-\alpha)^3$ & $\frac{\alpha(1152-2288\alpha+1552\alpha^2-377\alpha^3)}{1024}$	1 & $8\sqrt{\frac{16-48\alpha+54\alpha^2-22\alpha^3}{1152-2288\alpha+1552\alpha^2-377\alpha^3}}$

report the modular surface-code thresholds achieved with each. In Table IV, we summarize the hardware parameter sets considered.

To address this comparison, we first investigate which near-term hardware upgrade yields the largest threshold improvement and whether the thresholds saturate, as in fusion-based schemes. Finally, we determine the hardware requirements for near-term fault-tolerant modular demonstrations.

To ensure fair comparison with Ref. [6], we simulate the infidelity and success rates of all protocols, including the physical error probability p for circuit-level noise. Explicitly, this means that all communication qubit measurements and all quantum gates on communication and memory qubits have a failure rate of p. In Ref. [6], code thresholds were reported that ranged from 0.01% to 0.4% including fusion-based emission and scattering schemes. Here, we used $p = 0.1\%$ to have error contribution in the range where we expect the threshold to exist, to compare the output GHZ states from several protocols. For the results on the GHZ state quality, we use 10^4 shots for each data point for the average output, which results in tiny error zones as presented in the plots.

A. Decoherence and hardware parameter sets

Decoherence is a dominant error noise source, which becomes more crucial since the GHZ state generation is probabilistic and memory gates are slow, causing significant decoherence. We model the performance of protocols with active time-tracking decoherence channels for both communication and memory qubits. We use a representative range of coherence and gate timing parameters that correspond to "Set-3" in Ref. [6], as summarized in Table III. The quantities $T_{\text{link}}^{\text{dec}}$ and $T_{\text{idle}}^{\text{dec}}$ denote coherence times of the communication and memory qubits during entanglement attempts and idling, respectively. Operational times include t_{link} for one entanglement attempt and t_{meas} for communication-qubit measurement. Gate durations are given by $t_P^{\text{c/m}}$ (communication/memory) for single-qubit Pauli (P) gates ($P \in \{X, Y, Z, H\}$), and t_{CZ}, t_{CX}, t_{CY} for two-qubit gates on the memory, along with t_{SWAP} for the SWAP operation. While the simulation framework supports separate T_1 and T_2 coherence times for both communication and memory

TABLE III. Coherence and operation times for each module. The values are reported with respect to t_{link}, which is assumed to take one unit of time. For color centers in diamond, $t_{\text{link}} = 6 \times 10^{-6}$ s [6].

Coherence Time	Value
$T_{\text{link}}^{\text{dec}}$	10^6
$T_{\text{idle}}^{\text{dec}}$	10^6
Operation times	
T_{link}	1
T_{meas}	1
t_p^c	0.01
t_p^m	100
T_{CZ}, t_{CX}, t_{iCY}	100
T_{SWAP}	300

TABLE IV. Hardware improvement parameter sets for the single-shot emission scheme. The parameters are presented in a total of 18 sets, which show gradual improvement toward an ideal scenario. Set ES-1 corresponds to "Future parameters (FP)" considered in Ref. [6]. The sets trace a path from the near-term hardware baseline toward the near-ideal regime, serving as waypoints to quantify the hardware requirements for each architectural milestone.

Set	F_{prep}	P_{DE}	μ	η_{ph}
ES-1 (FP)	0.999	0.01	0.95	0.4474
ES-2	0.999	0.0	0.95	0.4474
ES-3	0.999	0.0	0.96	0.5
ES-4	0.999	0.0	0.97	0.6
ES-5	1	0.0	0.975	0.65
ES-6	1	0.0	0.98	0.7
ES-7	1	0.0	0.9825	0.75
ES-8	1	0.0	0.985	0.8
ES-9	1	0.0	0.9875	0.85
ES-10	1	0.0	0.99	0.9
ES-11	1	0.0	0.9925	0.95
ES-12	1	0.0	0.995	0.96
ES-13	1	0.0	0.9975	0.97
ES-14	1	0.0	0.998	0.98
ES-15	1	0.0	0.9985	0.985
ES-16	1	0.0	0.999	0.999
ES-17	1	0.0	1	0.999
ES-18	1	0.0	1	1

qubits, we assume $T^{\text{dec}}_{1,\text{link}} = T^{\text{dec}}_{2,\text{link}}$ and $T^{\text{dec}}_{1,\text{idle}} = T^{\text{dec}}_{2,\text{idle}}$ throughout, effectively modeling qubit noise as depolarizing. This also makes our surface code unbiased for the hardware errors, since there is no bias against X or Z errors from the hardware via the decoherence mechanism [6].

Next, we describe the emission-scheme hardware parameter sets considered for this work, as presented in Table IV. There are 18 emission-scheme (ES) parameter sets (ES-1 to ES-18) that show gradual improvement toward a noiseless scenario. The first set, ES-1, is the same as the future-parameter (FP) set for the emission-based scheme considered in Ref. [6]. From set ES-2, we apply our approximation model for the elimination of the double-excitation error under the state assumptions in Sec. II.

B. Bright state parameter influence

Given the full set of hardware, coherence, and gate timing parameters, the bright-state parameter α can be tuned to control both the success probability and the infidelity of the output GHZ state. As shown earlier, in the absence of noise, both quantities are functions of α: smaller (larger) values of α yield lower (higher) infidelity and success rates. Under realistic noise, these quantities must be computed explicitly for each protocol to determine the optimal α to determine exact values. As a starting point, we set $\alpha_{\text{base}} = \alpha_{\text{distil}} = \alpha$, for the base raw GHZ state and the distilling state, respectively. Following Ref. [6], we impose the thresholds $F_{\text{GHZ}} > 0.98$ and $P_{\text{succ}} > 10^{-3}$ to identify operating points relevant for toric surface code implementations. For the more general case $\alpha_{\text{base}} \neq \alpha_{\text{distil}}$, we present numerical performance of the distillation protocols in Appendix D when sweeping both α_{base}, and α_{distil}.

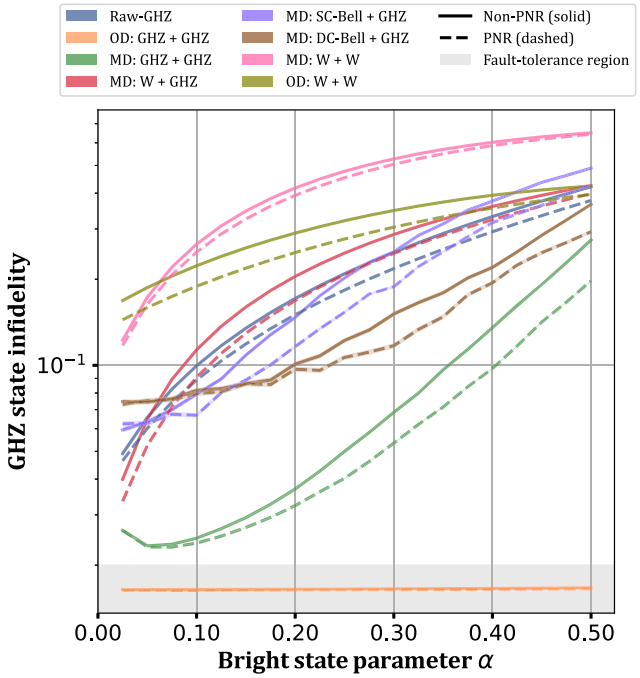


FIG. 6. GHZ state infidelity $(1 - F_{\text{GHZ}})$ for varying bright-state parameter (α) for ES-2 hardware parameter set, coherence times $T = 10^6$, and physical error rate of $p = 10^{-3}$.

The dependence of infidelity and success probability on α is shown in Figs. 6 and 7, respectively. These plots use the ES-2 hardware parameters. We observe that $P_{\text{succ}} > 10^{-3}$ is typically achieved for $\alpha > 0.1$ across most protocols. However, only the OD-GHZ protocol meets the fidelity threshold $F_{\text{GHZ}} > 0.98$. Since the fidelity of the OD-GHZ protocol is independent of α, we can choose $\alpha = 0.5$ to maximize its success rate. We now proceed to vary other hardware parameters to explore the full fault-tolerance regions of all protocols. For all other protocols, we choose the lowest value of α to be 0.025 to strongly bias for the lowest infidelity since that is the bottleneck in the performance based on Fig. 6. For these protocols, we need to lower the fidelity more, with hardware improvements. Unless stated otherwise, we use $\alpha = \alpha_{\text{base}} = \alpha_{\text{distil}} = 0.025$ for further analysis.

C. Varying the coherence times

In addition to the fixed coherence times listed in Table III, it is informative to evaluate protocol performance over a wider range of coherence values. We find numerically that

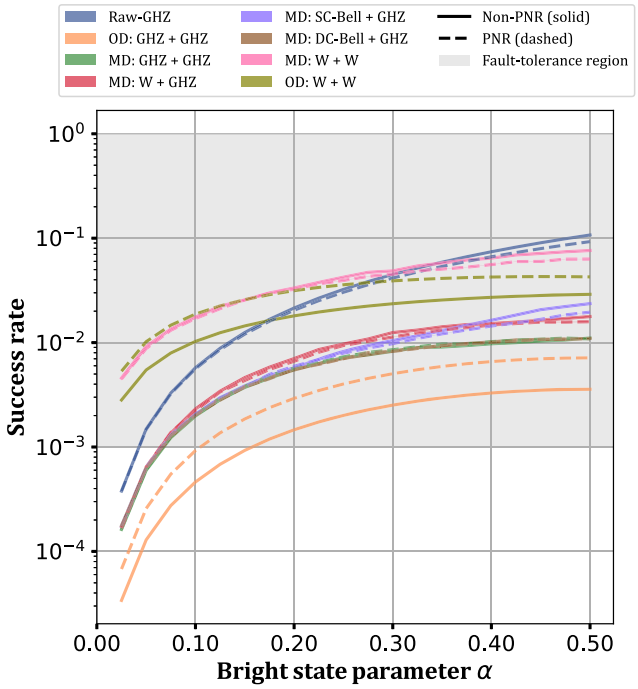


FIG. 7. Success rates for GHZ state (P_{succ}) for varying bright-state parameter (α) for ES-2 hardware parameter set, coherence times $T = 10^6$, and physical error rate of $p = 10^{-3}$.

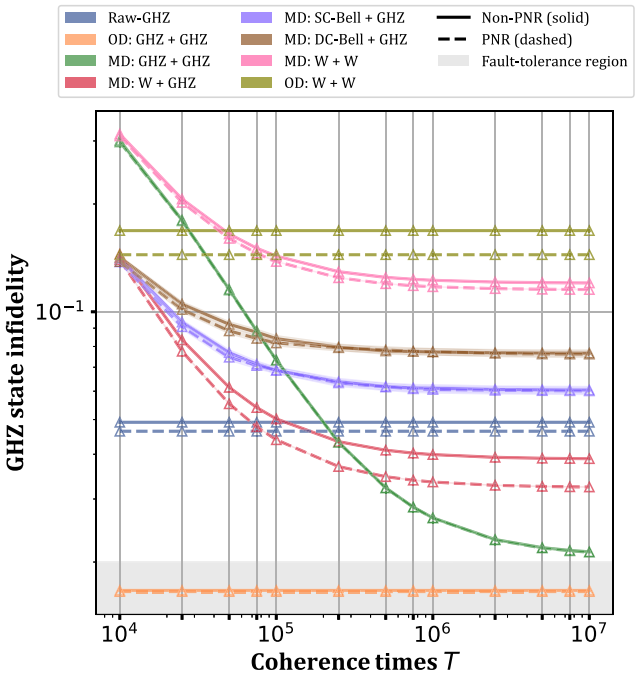


FIG. 8. GHZ state infidelity ($1 - F_{\text{GHZ}}$) for varying coherence times ($T = T_{\text{link}}^{\text{dec}} = T_{\text{idle}}^{\text{dec}}$) for the ES-2 hardware parameter set, physical error rate of $p = 10^{-3}$, tuned at $\alpha = 0.025$.

the success probabilities are largely insensitive to changes in coherence times. However, the GHZ state fidelity shows a strong dependence as presented in FIG. 8. We analyze infidelity as a function of coherence time T, sweeping from 10^4 to 10^7, which spans the parameter range defined by set-1 to set-3 in Ref. [6] and beyond.

Among the protocols, those involving memory-based distillation are most sensitive to short coherence times due to their reliance on two-qubit memory gates. For $T = 10^4$, these protocols fail to reach meaningful fidelities for quantum error correction. In contrast, setting $T = 10^6$ yields optimistic fidelity values, allowing us to identify fault-tolerant regions for distillation-based protocols.

D. Hardware parameters roadmap

For fixed choices of coherence times, physical error rate, and bright-state parameter, we now examine which hardware improvements are required to meet the fault-tolerance criteria for surface codes. This analysis is shown in Figs. 9 and 10, which display success probabilities and infidelities, respectively, based on the parameters presented in Table IV. Rather than serving as primary results, the parameter sets are presented as supporting evidence for three architectural conclusions established in previous subsections: the viability of gate-free optical distillation, the outsized impact of PNR detection, and the continuous scaling of thresholds. Among all protocols, the OD-GHZ protocol yields the lowest infidelity, and it meets the fault-tolerance requirements under hardware set ES-2 for both PNR and non-PNR detectors.

However, success probabilities show a substantial difference between the two detector types. In the PNR case, the OD-GHZ protocol achieves higher success rates compared to the non-PNR case. Acceptable success rates for the OD-GHZ protocol are first observed at hardware set ES-5. Memory distillation protocols begin to reach near-usable fidelities only at higher hardware quality but still below the desired threshold due to the gate noise.

The OD-W protocol shows better success rates than memory-based protocols but only becomes relevant under PNR detection and with significant hardware improvements. Among all considered approaches, the OD-GHZ protocol consistently performs best across hardware sets, with PNR detectors providing a clear advantage as parameters approach ideal values.

Interestingly, beyond the hardware set ES-14, the fidelity of the raw GHZ state becomes high enough that additional distillation is no longer needed. Nonetheless, the success rate for the raw GHZ protocol remains relatively low due to its dependence on small α and does not match the performance of the OD-GHZ protocol for equivalent fidelity. From the noiseless analysis of raw GHZ generation, we estimate that a value of $\alpha = 0.0762$ is needed to meet the surface code threshold, that is, a fidelity of at least 0.98. Now, in the presence of noise, the fidelity would drop due to noise, and a much lower value of α would be required to bring the fidelity to 0.98, in line with

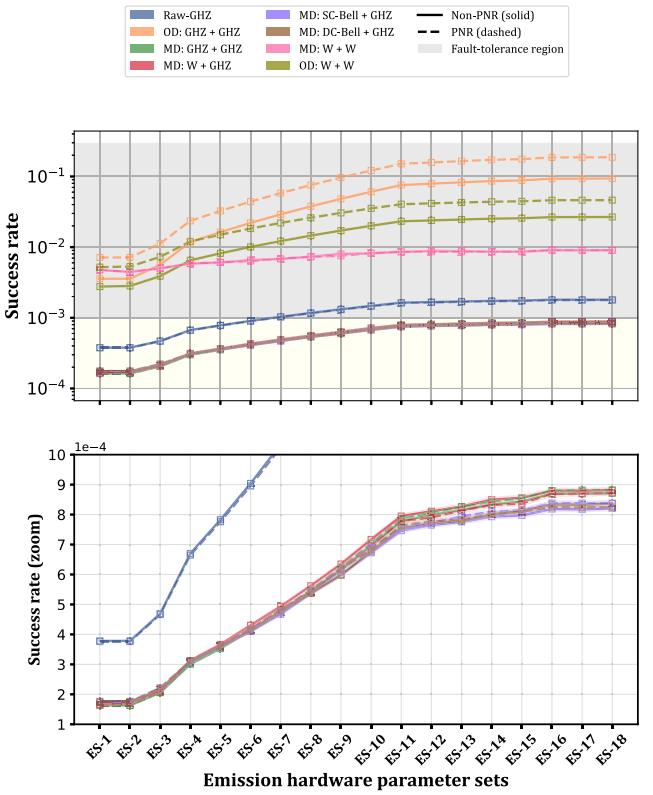


FIG. 9. GHZ success rates (P_{succ}) for varying hardware parameters from ES-1 to ES-18 for coherence times of $T = 10^6$, physical error rates of $p = 10^{-3}$ tuned at $\alpha = 0.025$ (except OD-GHZ that is biased for $\alpha = 0.5$).

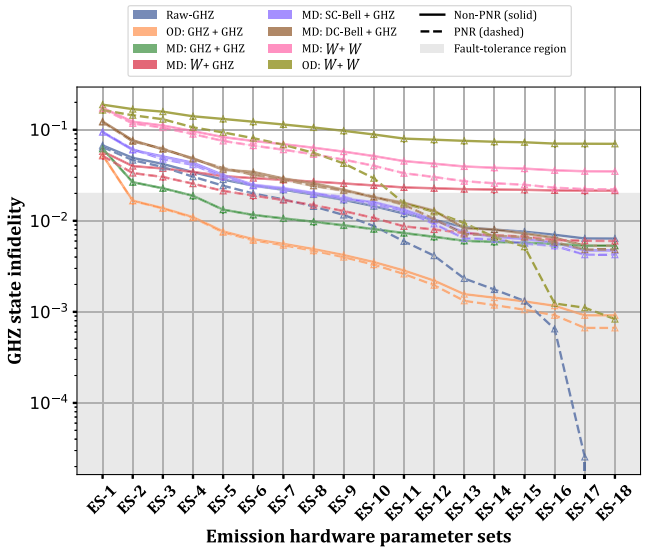


FIG. 10. GHZ state infidelity ($1 - F_{\text{GHZ}}$) for varying hardware parameters from ES-1 to ES-18 for coherence times of $T = 10^6$, physical error rates of $p = 10^{-3}$ tuned at $\alpha = 0.025$ (except OD-GHZ that is biased for $\alpha = 0.5$).

what we report. In Fig. 10, the infidelity for protocols other than the raw-GHZ state does not go to zero in the limit of noiseless hardware because there is a fixed physical error per gate and measurement of 10^{-3}, which is around the expected surface code threshold. In this analysis, we vary one parameter at a time to find fault-tolerance thresholds. To provide more insight useful to experimentalists for estimating the full hardware performance, we present two-dimensional plots of fault tolerance, sweeping both coherence times and hardware parameter sets in Appendix E.

E. Thresholds for fault-tolerance

Based on the hardware parameter plots (Figs. 9 and 10), we find the sets that yield code thresholds. These thresholds are calculated using the Union Find decoder due to its nearly linear time complexity [87]. The best fidelity protocol is OD-GHZ, and we search for its threshold performance starting with ES-1, for both PNR and non-PNR detector types. These GHZ states are input to the noisy stabilizer circuits shown in Fig. 1(b), and then QEC cycles are repeated to estimate the logical success (or error) rates. A detailed procedure of this calculation is presented in Appendix C. GHZ state generation in the distributed architecture is inherently probabilistic. To ensure synchrony in stabilizer measurements across the code lattice, we introduce a cutoff time, t_{cut}, which limits the maximum duration allowed for GHZ state generation before performing stabilizer measurements. A small t_{cut} reduces the available syndrome information, impairing error detection, while a large t_{cut} increases idle time, leading to greater decoherence. We therefore optimize t_{cut} to maximize the error-correction threshold of the surface code.

We evaluate the threshold performance by scanning different hardware sets, fixing the coherence time to $T = 10^6$ and setting the bright-state parameter to $\alpha = 0.5$, which is optimal for the OD-GHZ protocol. We simulate logical success rates by sweeping the physical error rate p (with $p = p_g = p_m$) for even code distances $d = 4, 6, 8, 10, 12$. The threshold extraction procedure is detailed in Appendix F. For PNR detectors, we obtain a circuit-level noise threshold of $p_{\text{th}} = 0.2498\% \pm 0.0023\%$ using hardware set ES-2. This exceeds the threshold of 0.13% reported in Ref. [6] for a Bell-pair fusion protocol with non-PNR detectors and ES-1 hardware (labeled FP in Ref. [6]). This improvement is attributed to reduced double-excitation errors and enhanced detection resolution in the single-shot emission-based hardware. The threshold curve is shown in Fig. 11. At the threshold value, the GHZ state has a success probability of $P_{\text{succ}} = 6.7 \times 10^{-5}$ and a fidelity of $F_{\text{GHZ}} = 0.9820$. The high threshold at ES-2 is enabled by the relatively high success rate with PNR detectors and the low impact of decoherence due to short GHZ generation times. This threshold competes with the threshold for direct scattering-based schemes studied in Ref. [6], which were typically reported to be around 0.35%.

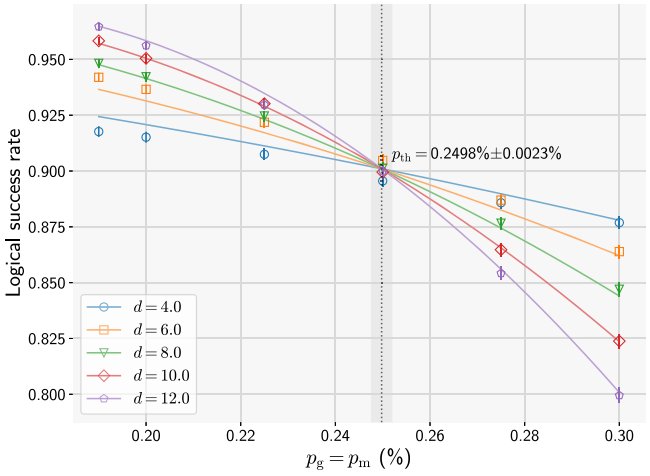


FIG. 11. Threshold plot for OD-GHZ protocol with PNR detectors for circuit-level noise. The threshold plot is evaluated for $T = 10^6$, and hardware parameter set ES-2, leading to a toric surface code threshold of $p_{\text{th}} = 0.2498\% \pm 0.0023\%$.

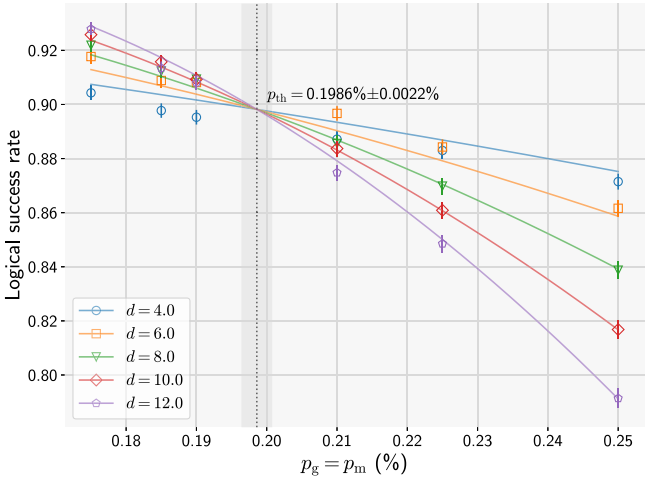


FIG. 12. Threshold plot for OD-GHZ protocol with non-PNR detectors for circuit-level noise. The threshold plot is evaluated for $T = 10^6$, and hardware parameter set ES-5, leading to a toric surface code threshold of $p_{\text{th}} = 0.1986\% \pm 0.0022\%$.

For non-PNR detectors as well, we are able to find a threshold with ES-5. A threshold of $p_{\text{th}} = 0.1986\% \pm 0.0022\%$ is found as shown in the error-rates plot in Fig. 12. At this threshold, the OD-GHZ protocol with non-PNR detectors yields a GHZ success probability of $P_{\text{succ}} = 3.38 \times 10^{-5}$ and fidelity $F_{\text{GHZ}} = 0.9824$. Because of the lower success rates in the non-PNR setting, additional hardware improvements are required to mitigate decoherence from longer GHZ generation times. Nevertheless, even with this modest hardware upgrade, the resulting threshold is sufficiently high and comparable to those achieved by fusion-based architectures reported in the literature [6,43]. This is due to the high fidelity that we obtain.

To draw a direct comparison with the results of Ref. [6], we can compare the stabilizer fidelity of hardware parameter set ES-1 (FP) of the previous best fusion-based scheme with the new OD-GHZ protocol with non-PNR detectors and PNR detectors, respectively. We do this since capturing the exact success rate of these complex scheduling fusion-based protocols is challenging, and stabilizer fidelity gives a good comparison of the input for threshold calculations. For the fusion-based protocol with $k = 11$ Bell pairs, the reported stabilizer fidelity was 0.9194 at a reported threshold of approximately 0.15%, and for the OD-GHZ protocol with PNR (non-PNR), we find a stabilizer fidelity of 0.9265 at $p_{\text{th}} = 0.2498\%$ (0.9255 at $p_{\text{th}} = 0.1986\%$).

Thus, we affirmatively answer the questions posed at the outset: the memory-gate bottleneck of fusion-based schemes is eliminated by OD-GHZ protocols without sacrificing thresholds. Among near-term upgrades, PNR detectors provide the largest gain, while requiring only modest optical improvements.

V. CONCLUSION

In this work, we investigated the advantages of emission-based single-shot GHZ entanglement generation with respect to fusion-based schemes for modular architectures. Building on experimentally demonstrated hardware, we proposed and analyzed multiple entanglement generation protocols capable of creating high-fidelity GHZ states across four modules, for each stabilizer of the code. These protocols fall into two categories: memory-based distillation and fully optical distillation. We benchmarked their performance in terms of success probability and output GHZ state fidelity under realistic noise and hardware parameters modeled after color centers in diamond.

Our results show that direct optical protocols, in particular the OD-GHZ protocol, can outperform memory-based protocols under modest hardware improvements, even in the absence of photon number-resolving detectors. One reason behind its performance is that it does not require two-qubit memory gates. Using realistic noise modeling and coherence constraints, we simulated distributed surface code operation and extracted fault-tolerance thresholds. Notably, we report a threshold of 0.24% with PNR detectors at hardware set ES-2 and a threshold of 0.19% for non-PNR detectors at ES-5. These values exceed or match previous results obtained with Bell-pair fusion-based approaches. We have shown that the thresholds with single-shot emission hardware can scale with improving hardware, unlike the previously evaluated fusion-based protocols that demonstrated maximum thresholds of approximately 0.16% restricted due to the slow two-qubit memory gates [6].

Overall, this work establishes the feasibility of implementing fault-tolerant surface codes in a distributed architecture using only demonstrated emission-based hardware.

Our proposed protocols enable early threshold crossings with modest hardware requirements and open pathways to scalable modular quantum computing architectures.

Several open directions follow from this work. An immediate extension is to study the performance of other topological or LDPC codes—such as hypergraph product codes or balanced product codes—within the same emission-based distributed architecture. These codes typically require higher-weight stabilizer measurements, which could benefit from generalizations of the single-shot GHZ generation protocols to larger module counts. Finally, incorporating realistic feedback and routing constraints into the protocol design at the logical level will be essential for bringing such modular architectures closer to practical fault tolerance.

ACKNOWLEDGEMENTS

We gratefully acknowledge support from the joint research program "Modular Quantum Computers" by Fujitsu Limited and Delft University of Technology, co-funded by the Netherlands Enterprise Agency under project number PPS2007, JST Moonshot R&D, Grants No. JPMJMS2061 and No. JPMJMS226C, and the Program for the Advancement of Next Generation Research Projects by Keio University. The authors thank SURF [88] for the support in using the Dutch National Supercomputer Snellius.

DATA AVAILABILITY

The data that support the findings of this article are openly available in reference [89] of the manuscript.

APPENDIX A: NOISE MODEL

1. Circuit-level noise

Circuit-level noise [90,91] is applied to state preparation, single- and two-qubit gates, and measurements. Note that the actual circuits we run only use the native gateset (as shown in Table III), including the single- or two-qubit gates for the communication and/or memory qubits, and measurements are only possible directly for communication qubits. State preparation, idling, and single-qubit gates are modeled as the ideal operation followed by a depolarizing channel with error probability p_g:

$$\mathcal{N}_{1q}(\rho) = (1-p_g)\rho + \frac{p_g}{3}\sum_{P_j\in\{X,Y,Z\}} P_j\rho P_j. \tag{A1}$$

Two-qubit gates are followed by a two-qubit depolarizing channel:

$$\mathcal{N}_{2q}(\rho) = (1-p_g)\rho + \frac{p_g}{15}\sum_{\substack{P_j,P_k\in\{I,X,Y,Z\}\\ \neq(I,I)}} (P_j\otimes P_k)\rho(P_j\otimes P_k). \tag{A2}$$

The measurement error probability p_m is modeled as the probability of flipping the measurement outcome. In simulations, we set $p_g = p_m = p$ and sweep p to characterize the effect of circuit-level noise.

We model decoherence as a continuous-time process using characteristic times T_1 (energy relaxation) and T_2 (pure dephasing). The probability to undergo decoherence over time t is $1-e^{-t/T^{\mathrm{dec}}}$. During entanglement generation attempts, memory qubits have coherence time $T^{\mathrm{dec}}_{\mathrm{link}}$, while in idle periods all qubits have $T^{\mathrm{dec}}_{\mathrm{idle}}$, with $T^{\mathrm{dec}}_{\mathrm{idle}} \geq T^{\mathrm{dec}}_{\mathrm{link}}$ for the emission-based hardware considered [92–94].

We implement decoherence channels in Kraus form:

$$\mathcal{N}^{\mathrm{dec}}(\rho) = \sum_{j=1}^{\kappa} K_j\rho K_j^\dagger, \qquad \sum_{j=1}^{\kappa} K_j^\dagger K_j = \mathbb{I}. \tag{A3}$$

The T_1 process is modeled by the generalized amplitude damping channel:

$$K_1^{\mathrm{GAD}} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & \sqrt{1-\gamma_1}\end{bmatrix}, \quad K_2^{\mathrm{GAD}} = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & \sqrt{\gamma_1}\\ 0 & 0\end{bmatrix},$$
$$K_3^{\mathrm{GAD}} = \frac{1}{\sqrt{2}}\begin{bmatrix}\sqrt{1-\gamma_1} & 0\\ 0 & 1\end{bmatrix}, \quad K_4^{\mathrm{GAD}} = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0\\ \sqrt{\gamma_1} & 0\end{bmatrix}, \tag{A4}$$

where $\gamma_1 = 1-e^{-t/T_1}$.

The T_2 process is modeled by phase damping:

$$K_1^{\mathrm{PD}} = \begin{bmatrix}1 & 0\\ 0 & \sqrt{1-\gamma_2}\end{bmatrix}, \qquad K_2^{\mathrm{PD}} = \begin{bmatrix}0 & 0\\ 0 & \sqrt{\gamma_2}\end{bmatrix}, \tag{A5}$$

where $\gamma_2 = 1-e^{-t/T_2}$.

When both channels act together, they yield an effective depolarizing process. All operations, including link generation, idling, and measurements, are assigned durations consistent with the main text, and decoherence is applied accordingly. In our simulations, this effective noise process is a time function and applied accordingly with the duration of the operations in action, in a fully time-scheduled modeling for all qubits.

2. Emission-based hardware noise

a. Bright-state parameter and state preparation

The bright-state parameter α characterizes the population of the communication qubit in the optically excited ("bright") state $|1\rangle$ relative to the ground ("dark") state $|0\rangle$ at the time of photon emission. The communication qubit, typically realized as the electronic spin of

a nitrogen-vacancy (NV^-) center in diamond [95], is initialized by optical pumping into the $|0\rangle$ state ($m_s = 0$), which fluoresces upon resonant excitation, in contrast to $|1\rangle$ ($m_s = -1$), which exhibits a substantial probability of nonradiative decay to a metastable state [96,97]. The parameter $\alpha \in [0, 1]$ is experimentally tunable by controlling the relative amplitude and duration of the optical or microwave pulses that prepare the initial superposition state

$$\sqrt{1-\alpha}\,|0\rangle + \sqrt{\alpha}\,|1\rangle. \tag{A6}$$

Upon application of a short, spin-selective resonant laser pulse (typically approximately 2 ns [98]) addressing the $|1\rangle \to |1_e\rangle$ transition, only the bright component $|1\rangle$ is excited to the optically excited state $|1_e\rangle$, which decays within $\tau_e \approx 20$ ns to emit a single photon. This process generates the emitter-photon entangled state

$$\sqrt{1-\alpha}\,|0\rangle|0_{\mathrm{ph}}\rangle + \sqrt{\alpha}\,|1\rangle|1_{\mathrm{ph}}\rangle, \tag{A7}$$

where $|0_{\mathrm{ph}}\rangle$ and $|1_{\mathrm{ph}}\rangle$ denote the zero- and one-photon Fock states, respectively.

Experimentally, the value of α plays a dual role:

1. *Impact on entanglement fidelity.* In the high-photon-loss regime ($\eta \ll 1$), the fidelity of the heralded two-qubit state with respect to the ideal maximally entangled state scales approximately linearly with α for low α [52]. At large α, fidelity drops sharply due to the increased relative weight of noise counts (dark counts, stray light, leakage from excitation pulses), as noise events falsely herald entanglement. Preparation errors in the unitary rotation or off-resonant excitations can exacerbate this drop.
2. *Impact on heralding probability.* The per-attempt success probability (for a single-click Bell pair) p_{click} scales linearly with α in the high-loss regime:

$$p_{\mathrm{succ}} \approx \eta\alpha. \tag{A8}$$

Imbalances in η can be compensated by adjusting α to maximize fidelity for a given p_{succ} [52].

In practice, α is optimized based on the specific trade-off between fidelity and rate for a given hardware configuration.

State-of-the-art NV center platforms report bright-state populations and preparation fidelities exceeding 99% [99,100], corresponding to state preparation and measurement errors below 10^{-3}. Such performance is essential for meeting the stringent error thresholds required for fault-tolerant operation of distributed quantum error-correction schemes.

The bright-state parameter is closely linked to the state-preparation fidelity F_{prep}, which characterizes the accuracy of preparing the intended qubit state before photon emission. Imperfect preparation of the desired superposition state (e.g., due to off-resonant excitation or control pulse errors) leads to phase uncertainty in the emitter state, which in our simulations is modeled as a pure dephasing channel acting on the qubit before photon emission:

$$\mathcal{E}_{\mathrm{prep}}(\rho) = F_{\mathrm{prep}}\rho + (1 - F_{\mathrm{prep}})Z\rho Z, \tag{A9}$$

This channel captures the reduction in emitter coherence due to imperfect initialization, mapping directly to the experimentally measured F_{prep} values. The channel is applied to the emitter density matrix ρ immediately after state preparation, prior to optical excitation.

Experimental reports of F_{prep} for NV-center platforms vary depending on the initialization protocol, optical pumping parameters, and environmental noise, with typical values in the range of 0.95–0.99 for optimized low-temperature setups [52,98,100]. In our numerical modeling, F_{prep} and α jointly determine the quality of the emitter-photon entangled state and, consequently, the performance of the overall entanglement-generation protocol.

b. Path-length difference

In an ideal entanglement-generation protocol, the Bell state $|\Psi^{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}$ is created with a well-defined relative phase between the two terms, depending on the detector click. In practice, optical path-length mismatches, residual frequency detuning between emitters, or imperfect phase stabilization introduce an additional phase ϕ in the state:

$$|\Psi_\phi\rangle = \frac{|01\rangle + e^{i\phi}|10\rangle}{\sqrt{2}}. \tag{A10}$$

If ϕ fluctuates between the attempts, the ensemble-averaged state acquires phase noise, which can be modeled as a dephasing process in the $\{|01\rangle, |10\rangle\}$ subspace. The amount of dephasing is determined by the variance σ_ϕ^2 of the phase instability. Following Refs. [43,101,102], the corresponding dephasing parameter is

$$\lambda = \frac{1}{2}\left[1 + \frac{I_1(\sigma_\phi^{-2})}{I_0(\sigma_\phi^{-2})}\right], \tag{A11}$$

where I_0 and I_1 are modified Bessel functions of the first kind of order 0 and 1, respectively. This yields an effective single-qubit dephasing channel acting on one of the modes (e.g., emitter A's photon) before detection:

$$\mathcal{E}_\phi(\rho) = \lambda\rho + (1 - \lambda)Z\rho Z. \tag{A12}$$

Generalization to a four-module setting: For networks with $N > 2$ modules, such as our four-module

architecture, each optical link (i,j) can have its own random phase offset ϕ_{ij} arising from independent path-length fluctuations and local oscillator phases. In the absence of active phase tracking, each ϕ_{ij} is drawn from a zero-mean distribution with variance σ_{ij}^2. The joint state ρ of the network then experiences correlated dephasing across all photon-pair subspaces:

$$\mathcal{E}_{\phi,\mathrm{multi}}(\rho) = \int \prod_{(i,j)} p_{ij}(\phi_{ij}) U_{\phi_{ij}}^{(i,j)} \rho\, U_{\phi_{ij}}^{(i,j)\dagger}\, \mathrm{d}\phi_{ij}, \quad \text{(A13)}$$

where $U_{\phi_{ij}}^{(i,j)}$ applies a phase rotation $e^{i\phi_{ij}}$ on the $|10\rangle$ component of the (i,j) subspace. If all links share the same source laser and stabilization system, these phases can be partially correlated (however, we can also consider them to be fully independent, just synced timely); a simplified model is then to assign a single Gaussian-distributed global phase ϕ_g plus independent link noise $\delta\phi_{ij}$.

Neglecting this effect in our simulations: In our current proposal, the dominant source of phase fluctuations is slow relative to the entanglement-generation time. Measured values for NV-based optical links give $\sigma_\phi \approx 14.3°$ [101], corresponding to $\lambda \approx 0.984$—a small reduction in off-diagonal coherence (for two-module setup), about approximately 0.16 dephasing factor. Given that this noise is much weaker than decoherence and hardware loss, we neglect this error source in our main simulations and assume phase tracking and corrections are available for small errors. We note that for long-term averaging or multiround protocols without active phase tracking, this noise channel would need to be explicitly included, and we leave this for a possible future research direction.

c. Double-excitation error

A critical nonideal process in emission-based entanglement generation is double optical excitation. This occurs when, during a finite-duration optical excitation pulse, an emitter is reexcited after emitting a photon, leading to the sequential emission of two photons in a single attempt. In a heralded entanglement protocol, where one photon detection is used to project two distant module emitters into an entangled emitter state, the undetected photon from a double excitation event leaks information to the environment, reducing the coherence of the spin-photon state and thereby lowering the fidelity of the heralded emitter-emitter entanglement [53,100,101].

Consider an emitter initialized in its ground state $|0\rangle$ and driven by a resonant optical π pulse. If the pulse duration is comparable to or longer than the excited-state lifetime T_1, there exists a finite probability that the emitter spontaneously emits a photon during the pulse and is then reexcited by the remaining optical field. This leads to a second emission event. In the high-loss regime typical of remote entanglement experiments, the probability of detecting both photons at the beam splitter is negligible; instead, one photon is typically lost, causing dephasing in the remaining spin-photon state.

Let p_{EE} denote the conditional probability that a given heralded detection is accompanied by an undetected extra photon emitted by the same emitter. Operationally, we define [103]:

$$p_{\mathrm{EE}} = \frac{P_2}{P_1 + P_2}, \quad \text{(A14)}$$

where P_1 is the probability of a single-photon emission and P_2 is the probability of a two-photon emission within one attempt, conditioned on at least one photon being detected in the heralding window. Since the detection efficiencies η_{Z} [for zero-phonon line (ZPL) photons] and η_{P} [for phonon-sideband (PSB) photons] differ, the experimentally accessible quantities are as follows:

$$P_2' = \eta_{\mathrm{P}}\eta_{\mathrm{Z}} P_2, \quad \text{(A15)}$$

$$P_1' = \eta_{\mathrm{Z}} P_1, \quad \text{(A16)}$$

leading to the experimentally useful expression [103]:

$$p_{\mathrm{EE}} = \frac{N_{\mathrm{coin}}}{\eta_{\mathrm{P}} N_{\mathrm{all}} + (1-\eta_{\mathrm{P}}) N_{\mathrm{coin}}}, \quad \text{(A17)}$$

where N_{coin} is the number of coincidences between ZPL and PSB detectors, and N_{all} is the total number of ZPL detection events in the heralding window over n repetitions.

Error channel representation: From the perspective of quantum channel modeling, an undetected photon corresponds to tracing over a photonic degree of freedom entangled with the spin. This yields a pure dephasing channel on the emitter-photon state:

$$\mathcal{E}_{\mathrm{EE}}(\rho) = (1 - p_{\mathrm{EE}})\rho + p_{\mathrm{EE}}\, Z\rho Z, \quad \text{(A18)}$$

where Z acts on the qubit basis $\{|0\rangle, |1\rangle\}$. This noise channel is applied immediately after the emission step in our superoperator-based simulations.

Model assumptions and mitigation: Reducing p_{EE} requires shortening the excitation pulse relative to T_1, thus minimizing the chance of reexcitation within the pulse duration. However, excessively short pulses can increase spectral bandwidth, leading to unwanted excitation of off-resonant transitions. Optimizing pulse length and shape is therefore essential to balance the entanglement generation rate and the final fidelity. While the double-excitation process described above can, in principle, degrade the fidelity of heralded states, in our simulations, we neglect this effect, consistent with the assumptions stated in the main text. Experimental studies of NV center remote entanglement report p_{EE} values in the range of

4%–7% [53,101]. More elaborate models, such as in Ref. [84], combine p_{EE} with off-resonant excitation probabilities into a single "excitation error" parameter p_{EE}, but for our purposes, p_{EE} alone captures the dominant contribution in the parameter regime relevant for our single-shot protocols. Following Ref. [83], we assume that the emitter is embedded in an optical cavity, the excitation pulse is shorter than the cavity-enhanced excited-state lifetime, and the pulse is optimally shaped such that nearby unwanted transitions undergo a complete Rabi cycle and return to the ground state [84,85]. Under these conditions, the probability of an undetected extra photon emission (p_{EE}) becomes negligible. This is supported by experimental and theoretical studies showing that, for NV centers under such optimized driving, both double-excitation and off-resonant excitation probabilities are strongly suppressed [104,105]. Therefore, in the parameter regime relevant to this work, p_{EE} can be safely ignored without impacting the accuracy of our quantum error-correction performance estimates.

d. Photon distinguishability

A key requirement for high-fidelity entanglement generation via photon interference is that the photons emitted by different nodes are "indistinguishable" in all degrees of freedom except the quantum information they encode. To describe and test the indistinguishability of two photons, one can make use of the so-called Hong-Ou-Mandel (HOM) effect [106]. If two perfectly indistinguishable single photons impinge simultaneously on the two input ports of a balanced beam splitter, the HOM effect dictates that they always exit through the same output port due to the two-photon quantum interference. This interference arises because the probability amplitudes for the two indistinguishable paths—both photons transmitted or both reflected—cancel exactly, resulting in photon bunching with zero coincidence probability. Any distinguishability—arising from differences in central frequency, spectral width, polarization, temporal shape, or arrival time—reduces the HOM visibility, which makes the HOM visibility a good measure for the pairwise photon distinguishability.

In our setup, photons from each node are collected into polarization-maintaining (PM) fibers and interfered on a fiber-coupled beam splitter. We assume that we align the polarizations to within better than 20 dB extinction [103], rendering polarization mismatch negligible. Spatial mode mismatch is also negligible due to the single-mode fiber coupling [103]. The dominant residual distinguishability is thus attributed to spectral and temporal variations only, in particular, fluctuating frequency offsets between photons from the two emitters.

Modeling indistinguishability: We model imperfect indistinguishability with a phenomenological "visibility" parameter $\mu_I \in [0,1]$, defined as the HOM two-photon interference visibility: $\mu_I = 1$ corresponds to perfect indistinguishability and $\mu_I = 0$ to completely distinguishable photons. When the photons have normalized spectral amplitudes $\psi(\omega)$ and $\phi(\omega)$, the indistinguishability can be expressed as the squared modulus of their spectral overlap [107,108]:

$$\mu_I = \left| \int_{-\infty}^{\infty} \psi^*(\omega)\phi(\omega)d\omega \right|^2. \tag{A19}$$

We adopt the term "intensity-definition" of indistinguishability (or visibility) to distinguish it from the amplitude-based definition introduced in Appendix B. The two notions are directly related via $\mu_I = |\mu|^2$ and exhibit equivalent noise sensitivity. In the main text, we consistently employ the intensity-based definition, as it corresponds to experimentally accessible measurements. Accordingly, all numerical values reported in the main text and used in simulations are based on this convention. For consistency with the POVM formalism, these values are internally converted to their square roots before being inserted into the POVM matrices, ensuring results remain directly comparable with experimental interpretations. In the case where both spectra are Gaussian with identical widths σ and a relative central frequency offset $\Delta\omega$, Eq. (A19) yields

$$\mu_I = \exp\left[-\frac{(\Delta\omega)^2}{4\sigma^2}\right], \tag{A20}$$

illustrating the exponential sensitivity of HOM visibility to spectral detuning.

In our simulations, μ_I directly scales the off-diagonal terms of the heralded two-qubit density matrix via the POVMs described in Appendix B, thereby capturing the reduction in coherence due to residual distinguishability. A similar approach has been used in prior NV-center entanglement experiments [53,97]. All photonic state projections that correspond to the desired detection patterns are calculated using the Kraus operators derived from these POVM elements.

Future improvements in μ_I could be achieved by stabilizing the emitter environment to suppress spectral diffusion, implementing active feedback to correct frequency drifts, or using cavity-enhanced emission to spectrally filter the photons [109]. Reaching $\mu_I \to 1$ is critical for scaling up remote entanglement protocols, as the entanglement fidelity scales linearly with μ_I in the absence of other dominant errors.

e. Effective photon detection efficiency

In any optical entanglement generation protocol, the probability that a resonantly emitted photon is actually detected is a critical figure of merit. We capture this in the "effective photon detection efficiency," denoted by η,

which represents the probability that an emitter prepared in the bright state $|1\rangle$ produces a zero-phonon-line (ZPL) photon that successfully propagates to the detector and triggers a click within the selected detection time window. Formally, η can be written as the product

$$\eta = p_{\mathrm{ZPL}}\, p_{\mathrm{col}}\, p_{\mathrm{trans}}\, p_{\mathrm{det}}\, p_{t_w}, \tag{A21}$$

where:

(i) p_{ZPL}: probability of emission into the ZPL rather than the PSB. For NV centers, $p_{\mathrm{ZPL}} \approx 0.03$ [110,111], with the remainder going into the PSB and thus being unsuitable for high-visibility interference.

(ii) p_{col}: collection efficiency of the optical interface (e.g., a lens or waveguide), including Fresnel losses at the diamond-air interface.

(iii) p_{trans}: transmission probability through the optical path to the beam splitter, given for fiber coupling by $p_{\mathrm{trans}} = 10^{-\ell_{\mathrm{fiber}}\gamma_{\mathrm{trans}}/10}$, where ℓ_{fiber} is the fiber length and γ_{trans} the attenuation coefficient in dB per unit length.

(iv) p_{det}: detector quantum efficiency at the ZPL wavelength.

(v) p_{t_w}: probability that the photon is emitted within the chosen time window t_w,

$$p_{t_w} = \int_{t_{\mathrm{start}}}^{t_{\mathrm{end}}} \frac{1}{\tau} e^{-t/\tau}\, dt = e^{-t_{\mathrm{start}}/\tau} - e^{-t_{\mathrm{end}}/\tau}, \tag{A22}$$

where τ is the excited-state lifetime.

In our simulations for short fiber links, p_{trans} is close to unity and can be neglected, but we retain it in the general expression (A21) for completeness.

Channel-based modeling: Photon loss and finite detection efficiency are modeled as an effective amplitude-damping (AD) channel acting independently on the photonic qubits emitted from each module. The Kraus operators for the AD channel with damping probability $1-\eta$ are

$$K_1^{\mathrm{AD}} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\eta} \end{bmatrix}, \qquad K_2^{\mathrm{AD}} = \begin{bmatrix} 0 & \sqrt{1-\eta} \\ 0 & 0 \end{bmatrix}. \tag{A23}$$

Applying these operators to the photon mode density matrices before measurement yields the state incorporating photon loss, which then feeds into the heralding and entanglement-generation calculations. This modeling is equivalent to assuming that any lost photon corresponds to an erasure of the corresponding detection event [112].

Outlook and mitigation strategies: Photon loss directly lowers the entanglement success probability without affecting the fidelity of successful events (since they are postselected). Improving η is therefore a central target for experimental optimization. Several strategies are available:

(i) Enhancing p_{ZPL} via resonant cavities or strain engineering to suppress PSB emission [109,113].

(ii) Increasing p_{col} using high numerical aperture optics, solid-immersion lenses, or integrated photonic structures [114].

(iii) Reducing transmission losses p_{trans} by minimizing fiber lengths, using low-loss fibers, or placing beam splitters closer to the emitters.

(iv) Employing detectors with higher quantum efficiency p_{det} and lower dark counts [115].

(v) Optimizing p_{t_w} by matching the detection window to the emitter lifetime while rejecting excitation-laser leakage.

With state-of-the-art techniques, single-NV η values exceeding a few percent are achievable, and cavity-enhanced platforms have demonstrated order-of-magnitude improvements [109].

f. (Non) photon number resolving (PNR) detectors

A PNR detector can discriminate between different numbers of incident photons within a given detection window, outputting not just a binary click or no-click event but an integer photon count. In contrast, a non-PNR detector only registers whether at least one photon arrived, producing a click regardless of whether one or multiple photons were incident [116].

In the context of entanglement-generation protocols using single-photon interference, the distinction between PNR and non-PNR detection directly affects both the success rates and the output fidelity. In non-PNR detection, multiphoton events can lead to false positives—cases where the heralding signal indicates successful entanglement, but the resulting state fidelity is degraded due to contributions from unwanted higher-photon-number components. With PNR capability, such events can be rejected by conditioning only on single-photon detection outcomes, thereby improving the fidelity at the expense of a reduced success rate (this is how we improved the success rates of the OD-GHZ protocol by making use of extra patterns using PNR configuration). The optimal choice depends on the photon statistics of the source, the target fidelity, and the desired entanglement rate.

POVM-based modeling: We model the detection process using a set of POVM elements $\{E_n\}$, where E_n corresponds to the projection onto an n-photon detection event at a given detector. Here, E_n can be interpreted as a detection pattern with a total number of photons n desired in the pattern. See Appendix B for details on this. These satisfy the completeness relation

$$\sum_{n=0}^{\infty} E_n = \mathbb{I}. \tag{A24}$$

In the non-PNR case, the effective "click" POVM is given by the sum of all nonzero photon-number outcomes:

$$E_{\text{click}}^{(\text{non-PNR})} = \sum_{n\geq 1} E_n. \tag{A25}$$

This models the fact that the detector cannot distinguish between one and multiple photons—all such cases are treated identically in the analysis.

For an ideal PNR detector, the POVM element corresponding to the detection of exactly n photons is simply E_n, corresponding to the exact target pattern. When the protocol requires conditioning on single-photon detection events, postselection is performed on $n = 1$ outcomes only, i.e.,

$$E_{\text{click}}^{(\text{PNR},n=1)} = E_1. \tag{A26}$$

In a realistic detector, optical loss and spurious counts must be taken into account. Losses are characterized by detection efficiency η_{det}, while dark counts occur with probability p_{dc} per detection window. These imperfections transform the ideal POVM elements Π_n into effective ones, $\tilde{\Pi}_n(\eta_{\text{det}}, p_{\text{dc}})$ [117]. For example, the effective "single-photon" POVM becomes a weighted mixture of the ideal Π_0, Π_1, and higher-order elements, with weights determined by η_{det} and p_{dc}. However, for the scope of this work, we consider all the detectors to be identical with no dark counts. The effects of η_{det} are fully absorbed in η.

While most quantum network experiments to date have relied on non-PNR detectors—typically avalanche photodiodes or superconducting nanowire single-photon detectors (SNSPDs)—significant progress has been made toward high-efficiency PNR detection. Several experimental devices exist for this research area, such as transition-edge sensors [118], multiplexed SNSPDs with high timing resolution and lower dark counts [119], and superconducting kinetic-inductance detectors enabling scalable PNR systems [120]. For quantum repeater and distributed computing architectures, PNR detectors enhance heralding fidelity in multiphoton interference and reduce background effects.

APPENDIX B: POVM CALCULATION FOR SINGLE-SHOT MULTI-PARTITE ENTANGLEMENT GENERATION

In this appendix, we present the derivation of POVMs relevant for the heralded generation of W, GHZ, and Bell states, following the framework of Ref. [108]. The construction is carried out on the basis of input photon Fock states and explicitly incorporates temporal-mode mismatch as well as multiphoton contributions arising in threshold detection. These POVMs serve as an analytical tool for modeling detector responses to specific click patterns, which, in turn, underpin the elementary-state generation protocols for the emitters, as discussed in the main text.

1. General framework of derivation

As described in the main text, the measurement module consists of the 4-by-4 beam-splitter network and four photon detectors in the output ports. The projection operator for n_k photons at the kth output port is defined as

$$|n_k\rangle_k\langle n_k| = \frac{1}{n_k!}\hat{r}_k^{\dagger n_k}|0\rangle_k\langle 0|\hat{r}_k^{n_k}, \tag{B1}$$

where $\hat{r}_k^\dagger$ and $\hat{r}_k$ are creation and annihilation operators at the kth port. The n-photon projector at the kth port is transformed into a projector for input photons using the beam-splitter transformation given in Eq. (9):

$$\begin{pmatrix} \hat{p}_0^\dagger \\ \hat{p}_1^\dagger \\ \hat{p}_2^\dagger \\ \hat{p}_3^\dagger \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} \hat{r}_0^\dagger \\ \hat{r}_1^\dagger \\ \hat{r}_2^\dagger \\ \hat{r}_3^\dagger \end{pmatrix}, \tag{B2}$$

where $\hat{p}_j^\dagger$ is the creation operator at the jth input mode. Note that, for mathematical simplicity, the subscripts in this appendix start from 0, whereas they start from 1 in the main text. For simplicity, we denote the beam-splitter transformation by $\hat{V}$. Then, the projector for the input modes is

$$|n_k\rangle_k\langle n_k| \xrightarrow{\hat{V}} \frac{1}{n_k!}\left(\sum_j V_{jk}\hat{p}_j^\dagger\right)^{n_k}|0\rangle_{\text{in}}\langle 0|\left(\sum_j V_{jk}\hat{p}_j\right)^{n_k}, \tag{B3}$$

where V_{jk} denotes the (j,k)-element of $\hat{V}$, and $|0\rangle_{\text{in}}\langle 0|$ represents the vacuum state of all input modes.

Since there are four output ports, the projection operator for all output modes is given by taking the tensor product of the single-port projection operator $|n_k\rangle_k\langle n_k|$ in Eq. (B1), for $k = 0, \ldots, 3$:

$$\bigotimes_{k=0}^{3}|n_k\rangle_k\langle n_k| = \bigotimes_{k=0}^{3}\left[\frac{1}{n_k!}\hat{r}_k^{\dagger n_k}|0\rangle_k\langle 0|\hat{r}_k^{n_k}\right]. \tag{B4}$$

Then, the projection operator for input states is given by

$$\bigotimes_{k=0}^{3}\frac{1}{n_k!}\left(\sum_j V_{jk}\hat{p}_j^\dagger\right)^{n_k}|0\rangle_{\text{in}}\langle 0|\left(\sum_j V_{jk}\hat{p}_j\right)^{n_k}. \tag{B5}$$

Hereafter, we incorporate temporal modes into the above discussion. When the photon detectors have a flat spectral response, the n_k photon projector at the kth detector is defined as

$$\hat{\mathcal{P}}_k^{(n_k)} = \frac{1}{n_k!}\int d^{n_k}\mathbf{\Omega}_k \bigotimes_{l=0}^{n_k-1}[\hat{r}_k^\dagger(\Omega_{kl})|0\rangle_k\langle 0|\hat{r}_k(\Omega_{kl})], \tag{B6}$$

where $\boldsymbol{\Omega}_k$ is a vector of length n_k, and Ω_{kl} is a lth element in the vector $\boldsymbol{\Omega}_k$. The creation and annihilation operators for the frequency mode Ω_{kl} satisfy the following commutation relations:

$$[\hat{r}(\Omega_{kl}), \hat{r}^\dagger(\Omega'_{kl})] = \delta(\Omega_{kl} - \Omega'_{kl}),$$
$$[\hat{r}(\Omega_{kl}), \hat{r}(\Omega'_{kl})] = 0. \tag{B7}$$

The projection operator corresponding to the detection of $\boldsymbol{n} = [n_0, \ldots, n_3]$ photons in the four detectors is defined as

$$\hat{\mathcal{P}}_{\boldsymbol{n}} = \bigotimes_{k=0}^{3} \hat{\mathcal{P}}_k^{(n_k)}. \tag{B8}$$

Next, we consider the temporal modes of the input photons. When all input photons have a single frequency, the input Hilbert space is spanned by the following basis:

$$|0000\rangle_{\text{in}}, |0001\rangle_{\text{in}}, |0010\rangle_{\text{in}}, \cdots, |1111\rangle_{\text{in}}. \tag{B9}$$

For simplicity, we represent these input states using a bit string $\boldsymbol{m} = m_0 m_1 m_2 m_3$ as

$$|\boldsymbol{m}\rangle_{\text{in}} = |m_0 m_1 m_2 m_3\rangle_{\text{in}} \tag{B10}$$

$$= \bigotimes_{j=0}^{3} (\hat{p}_j^\dagger)^{m_j} |0\rangle_{\text{in}}, \tag{B11}$$

where $m_i = 0, 1$ for $i = 0, \ldots, 3$. By incorporating the temporal mode function $\phi_j(\omega)$ and the phase shift $e^{i\omega\tau_j}$ into the input states, the creation operator at the jth input mode is expressed as

$$\hat{p}_j^\dagger \to \int d\omega\, \phi_j^*(\omega) e^{i\omega\tau_j} \hat{p}_j^\dagger(\omega). \tag{B12}$$

Furthermore, the input annihilation operator is transformed by the beam-splitter matrix in Eq. (B2) as

$$\int d\omega\, \phi_j^*(\omega) e^{i\omega\tau_j} \hat{p}_j^\dagger(\omega) \xrightarrow{\hat{V}} \sum_{k=0}^{3} V_{jk} \hat{R}_{jk}^\dagger, \tag{B13}$$

where $\hat{R}_{jk}^\dagger$ is the creation operator at the kth output port associated with the temporal mode of the jth input port:

$$\hat{R}_{jk}^\dagger = \int d\omega\, \phi_j^*(\omega) e^{-i\omega\tau_j} \hat{r}_k^\dagger(\omega). \tag{B14}$$

To construct the POVMs on the input space spanned by $|\boldsymbol{m}\rangle_{\text{in}}$, we introduce the isometry $\hat{\mathcal{V}}$ defined as

$$|\boldsymbol{m}\rangle_{\text{in}} = \bigotimes_{j=0}^{3} (\hat{p}_j^\dagger)^{m_j} |0\rangle_{\text{in}}$$
$$\xrightarrow{\hat{\mathcal{V}}} \bigotimes_{j=0}^{3} \left(\sum_{k=0}^{3} V_{jk} \hat{R}_{jk}^\dagger \right)^{m_j} |0\rangle_{\text{out}} \tag{B15}$$

$$=: |\hat{\mathcal{V}}\boldsymbol{m}\rangle_{\text{out}}, \tag{B16}$$

which is given by

$$\hat{\mathcal{V}} = \sum_{\boldsymbol{m}} \bigotimes_{j=0}^{3} \left(\sum_{k=0}^{3} V_{jk} \hat{R}_{jk}^\dagger \right)^{m_j} |0\rangle_{\text{out}} \langle \boldsymbol{m}|_{\text{in}}. \tag{B17}$$

Using this isometry, we obtain the POVMs acting on the input photons and incorporating temporal-mode mismatch effects:

$$\hat{E}_{\boldsymbol{n}} = \hat{\mathcal{V}}^\dagger \hat{\mathcal{P}}_{\boldsymbol{n}} \hat{\mathcal{V}}$$
$$= \sum_{\boldsymbol{m},\boldsymbol{m}'} |\boldsymbol{m}\rangle_{\text{in}} \langle \boldsymbol{m}'|$$
$$\times \int d^{N_{\text{in}}}\boldsymbol{\Omega}\, M_{\boldsymbol{m}\boldsymbol{n}}(\boldsymbol{\Omega}) M^*_{\boldsymbol{m}'\boldsymbol{n}}(\boldsymbol{\Omega}), \tag{B18}$$

where we simply wrote the multiple integral over frequencies $\boldsymbol{\Omega}$ corresponding to N_{in} detected photons, and

$$M_{\boldsymbol{m}\boldsymbol{n}}(\boldsymbol{\Omega}) = {}_{\text{out}}\langle \hat{\mathcal{V}}\boldsymbol{m}| \bigotimes_{k'=0}^{3} \left[\bigotimes_{l=0}^{n_{k'}-1} \frac{1}{\sqrt{n_{k'}!}} \hat{r}_{k'}^\dagger(\Omega_{k'l}) \right] |0\rangle_{\text{out}}. \tag{B19}$$

When the vacuum state is injected in the jth input port, the relevant terms in Eq. (B15) become an identity. Then, we can simplify $E_{\boldsymbol{n}}$ by removing the terms corresponding to the vacuum inputs. We introduce the new indices $\boldsymbol{j}$ representing the input ports of injected single photons, defined as

$$\boldsymbol{j} = \{j \in \{0, 1, 2, 3\} | m_j \neq 0\}. \tag{B20}$$

Since the bit string $\boldsymbol{m}$ representing the input photon number is in one-to-one correspondence with the input port indices $\boldsymbol{j}$, the input state vector $|\boldsymbol{m}\rangle_{\text{in}}$ can be replaced with $|\boldsymbol{j}\rangle$. For example, $|\boldsymbol{m}\rangle_{\text{in}} = |0110\rangle_{\text{in}}$ leads to $|\boldsymbol{j}\rangle = |1, 2\rangle$. Then, the simplified POVMs are given by

$$\hat{E}_{\boldsymbol{n}} = \sum_{\boldsymbol{j},\boldsymbol{j}'} |\boldsymbol{j}\rangle_{\text{in}} \langle \boldsymbol{j}'| \int d^{N_{\text{in}}}\boldsymbol{\Omega} M_{\boldsymbol{j}\boldsymbol{n}}(\boldsymbol{\Omega}) M^*_{\boldsymbol{j}'\boldsymbol{n}}(\boldsymbol{\Omega}), \tag{B21}$$

where

$$M_{\boldsymbol{jn}}(\boldsymbol{\Omega}) = {}_{\text{out}}\langle 0| \bigotimes_{j\in\boldsymbol{j}} \left(\sum_{k=0}^{3} V_{jk}\hat{R}_{jk} \right) \bigotimes_{k'=0}^{3} \left[\bigotimes_{l=0}^{n_{k'}-1} \frac{1}{\sqrt{n_{k'}!}} \hat{r}_{k'}^{\dagger}(\Omega_{k'l}) \right] |0\rangle_{\text{out}}. \tag{B22}$$

2. POVM for W states

As described in the main text, the measurement module consists of the 4-by-4 beam-splitter network and four photon detectors, and it heralds W-state generation when a single photon is detected in one of the four detectors [77]. In this appendix, we consider the case where the photon clicks in the first detector, which generates the following four-partite W state:

$$|W_4\rangle = \frac{1}{2}(|1000\rangle + |0100\rangle + |0010\rangle + |0001\rangle). \tag{B23}$$

When the threshold detectors are used, it is impossible to distinguish between single-photon detection and multi-photon detection. Since at most four photons are injected, the number of detected photons is also at most four. Accordingly, the following four projectors capture all output patterns associated with $|W\rangle$ generation:

$$\hat{\mathcal{P}}_{1000} = \int d\Omega_0\, \hat{r}_0^{\dagger}(\Omega_0)|0\rangle_0\langle 0|\hat{r}_0(\Omega_0), \tag{B24}$$

$$\mathcal{P}_{2000} = \frac{1}{2}\iint d\Omega_0 d\Omega_1 \times \hat{r}_0^{\dagger}(\Omega_0)\hat{r}_0^{\dagger}(\Omega_1)|0\rangle_0\langle 0|\hat{r}_0(\Omega_0)\hat{r}_0(\Omega_1), \tag{B25}$$

$$\hat{\mathcal{P}}_{3000} = \frac{1}{6}\int d\Omega_0 d\Omega_1 d\Omega_2 \bigotimes_{l=0}^{2} \hat{r}_0^{\dagger}(\Omega_l)|0\rangle_0\langle 0| \bigotimes_{l=0}^{2} \hat{r}_0(\Omega_l), \tag{B26}$$

$$\hat{\mathcal{P}}_{4000} = \frac{1}{24}\int d\Omega_0 d\Omega_1 d\Omega_2\, d\Omega_3 \bigotimes_{l=0}^{3} \hat{r}_0^{\dagger}(\Omega_l)|0\rangle\langle 0| \bigotimes_{l=0}^{3} \hat{r}_0(\Omega_l). \tag{B27}$$

We derive the POVM operators $\hat{E}_{1000}$, $\hat{E}_{2000}$, $\hat{E}_{3000}$, and $\hat{E}_{4000}$ in the following subsections.

a. $\hat{E}_{1000}$

We derive the POVM $\hat{E}_{1000}$ corresponding to the projector $\hat{\mathcal{P}}_{1000}$, i.e., the case where $n_0 = 1$ and $n_1 = n_2 = n_3 = 0$ in Eq. (B22). In this case, only one photon is injected into the central entangling module, represented by the following state vectors:

$$|\boldsymbol{j}\rangle = |j\rangle \in \{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}. \tag{B28}$$

Then, the elements $M_{\boldsymbol{j}1000}$ of the POVM E_{1000} are given by

$$M_{\boldsymbol{j}1000}(\boldsymbol{\Omega}) = {}_{\text{out}}\langle 0|V_{j0}\hat{R}_{j0}\hat{r}_0^{\dagger}(\boldsymbol{\Omega})|0\rangle_{\text{out}}. \tag{B29}$$

As shown in Eq. (B2), all elements in the 0th column of the beam-splitter matrix V_{j0} are equal to $1/2$. Using the commutation relation given in Eq. (B7),

$$\begin{aligned} M_{j1000}(\Omega) &= \frac{1}{2}{}_{\text{out}}\langle 0|\hat{R}_{j0}\hat{r}_0^{\dagger}(\Omega)|0\rangle_{\text{out}} \\ &= \frac{1}{2}{}_{\text{out}}\langle 0| \int d\omega\, \phi_j^*(\omega)e^{i\omega\tau_j}\hat{r}_0(\omega)\hat{r}_0^{\dagger}(\Omega)|0\rangle_{\text{out}} \\ &= \frac{1}{2}\phi_j^*(\Omega)e^{i\Omega\tau_j}. \end{aligned} \tag{B30}$$

Substituting Eq. (B30) into Eq. (B21), we obtain

$$\hat{E}_{1000} = \frac{1}{4}\sum_{jj'}|\boldsymbol{j}\rangle_{\text{in}}\langle \boldsymbol{j}'| \times \int d\Omega\, \phi_j^*(\Omega)\phi_{j'}(\Omega)e^{-i\Omega(\tau_{j'}-\tau_j)}, \tag{B31}$$

which is simplified to

$$\hat{E}_{1000} = \frac{1}{4}\sum_{jj'}\mu_{jj'}|\boldsymbol{j}\rangle_{\text{in}}\langle \boldsymbol{j}'|, \tag{B32}$$

where $|\boldsymbol{j}\rangle$ and $|\boldsymbol{j}'\rangle$ belong to the same set in Eq. (B28), and $\mu_{jj'}$ denotes the visibility parameter between the photons from the jth and j'th input ports:

$$\mu_{jj'} = \int d\Omega\, \phi_j^*(\Omega)\phi_{j'}(\Omega)e^{-i\Omega(\tau_{j'}-\tau_j)}. \tag{B33}$$

We refer to this as the "amplitude definition" of indistinguishability or visibility. This should not be confused with the "squared" definition of indistinguishability as defined in Appendix A 2 d, where we define it based on intensity and is the one that is considered for the numerical simulations in the main text.

b. $\hat{E}_{2000}$

As in the previous subsection, when two photons are detected at the first detector, the photon-number indices are $n_0 = 2$ and $n_1 = n_2 = n_3 = 0$ in Eq. (B22). In this case, the corresponding input states are given by

$$|\boldsymbol{j}\rangle = |j_0 j_1\rangle \in \{|01\rangle, |02\rangle, |03\rangle, \times |12\rangle, |13\rangle, |23\rangle\}. \tag{B34}$$

The elements M_{j2000} of the POVM E_{2000} are given by

$$M_{j2000}(\boldsymbol{\Omega}) = \frac{1}{4\sqrt{2}}\,{}_{\text{out}}\langle 0|\hat{R}_{j_0 0}\hat{R}_{j_1 0}\hat{r}_0^\dagger(\Omega_0)\hat{r}_0^\dagger(\Omega_1)|0\rangle_{\text{out}}. \tag{B35}$$

Using the commutation relation in Eq. (B7), we obtain

$$\begin{aligned}&{}_{\text{out}}\langle 0|\hat{r}(\omega_0)\hat{r}(\omega_1)\hat{r}^\dagger(\Omega_0)\hat{r}^\dagger(\Omega_1)|0\rangle_{\text{out}}\\ &= \delta(\omega_0-\Omega_0)\delta(\omega_1-\Omega_1)\\ &\quad + \delta(\omega_0-\Omega_1)\delta(\omega_1-\Omega_0),\end{aligned} \tag{B36}$$

and we find

$$\begin{aligned}&M_{j2000}(\boldsymbol{\Omega})\\ &= \frac{1}{4\sqrt{2}}\{\phi^*_{j_0}(\Omega_0)\phi^*_{j_1}(\Omega_1)\exp\left[i(\Omega_0\tau_{j_0}+\Omega_1\tau_{j_1})\right]\\ &\quad + \phi^*_{j_1}(\Omega_0)\phi^*_{j_0}(\Omega_1)\exp\left[i(\Omega_1\tau_{j_0}+\Omega_0\tau_{j_1})\right]\}\\ &= \frac{1}{4\sqrt{2}}\sum_{\sigma\in S_2}\prod_{k=0}^{1}\phi^*_{j_{\sigma(k)}}(\Omega_k)\exp(i\Omega_k\tau_{j_{\sigma(k)}}).\end{aligned} \tag{B37}$$

Here, σ is a permutation on two elements, and S_2 is the symmetric group on the set $\{0,1\}$. Therefore, the POVM operator $\hat{E}_{2000}$ is given by

$$\hat{E}_{2000} = \frac{1}{32}\sum_{\boldsymbol{j}\boldsymbol{j}'}\left[\sum_{\sigma,\sigma'\in S_2}\prod_{k=0}^{1}\mu_{j_{\sigma(k)}j'_{\sigma'(k)}}\right]|\boldsymbol{j}\rangle_{\text{in}}\langle\boldsymbol{j}'|, \tag{B38}$$

where $|\boldsymbol{j}\rangle$ and $|\boldsymbol{j}'\rangle$ belong to the same set in Eq. (B34).

c. $\hat{E}_{3000}$

When three photons are detected at the first detector, the photon-number indices are $n_0 = 3$ and $n_1 = n_2 = n_3 = 0$ in Eq. (B22). In this case, photons are injected into the three input ports, represented by

$$|\boldsymbol{j}\rangle = |j_0 j_1 j_2\rangle \in \{|012\rangle, |013\rangle, |023\rangle, |123\rangle\}. \tag{B39}$$

The elements M_{j3000} of POVM $\hat{E}_{3000}$ are given by

$$M_{j3000}(\boldsymbol{\Omega}) = \frac{1}{8\sqrt{6}}\,{}_{\text{out}}\langle 0|\bigotimes_{k=0}^{2}\hat{R}_{j_k 0}\hat{r}_0^\dagger(\Omega_k)|0\rangle_{\text{out}}. \tag{B40}$$

Using the commutation relation in Eq. (B7), we obtain

$${}_{\text{out}}\langle 0|\bigotimes_{k=0}^{2}\hat{r}_0(\omega_k)\hat{r}_0^\dagger(\Omega_k)|0\rangle_{\text{out}} = \sum_{\sigma\in S_3}\prod_{k=0}^{2}\delta(\omega_k-\Omega_{\sigma(k)}), \tag{B41}$$

and we find

$$M_{j3000}(\boldsymbol{\Omega}) = \frac{1}{8\sqrt{6}}\sum_{\sigma\in S_3}\prod_{k=0}^{2}\phi^*_{j_{\sigma(k)}}(\Omega_k)\exp(i\Omega_k\tau_{j_{\sigma(k)}}), \tag{B42}$$

where S_3 denotes the symmetric group on three elements, namely $\{0,1,2\}$. Therefore, the POVM operator $\hat{E}_{3000}$ is given by

$$\hat{E}_{3000} = \frac{1}{384}\sum_{\boldsymbol{j}\boldsymbol{j}'}\left[\sum_{\sigma,\sigma'\in S_3}\prod_{k=0}^{2}\mu_{j_{\sigma(k)}j'_{\sigma'(k)}}\right]|\boldsymbol{j}\rangle_{\text{in}}\langle\boldsymbol{j}'|, \tag{B43}$$

where $|\boldsymbol{j}\rangle$ and $|\boldsymbol{j}'\rangle$ belong to the same set in Eq. (B39).

d. $\hat{E}_{4000}$

When four photons are detected at the first detector, the photon-number indices are $n_0 = 4$ and $n_1 = n_2 = n_3 = 0$ in Eq. (B22). In this case, the photons are injected into all four input ports, represented by the state vector $|\boldsymbol{j}\rangle = |0123\rangle$. Then, the corresponding elements M_{j4000} of POVM $\hat{E}_{4000}$ are given by

$$M_{j4000}(\boldsymbol{\Omega}) = \frac{1}{16\sqrt{4!}}\,{}_{\text{out}}\langle 0|\bigotimes_{k=0}^{3}\hat{R}_{k0}\hat{r}_0^\dagger(\Omega_k)|0\rangle_{\text{out}}. \tag{B44}$$

Using the commutation relation in Eq. (B7), we obtain

$${}_{\text{out}}\langle 0|\bigotimes_{k=0}^{3}\hat{r}_0(\omega_k)\hat{r}_0^\dagger(\Omega_k)|0\rangle_{\text{out}} = \sum_{\sigma\in S_4}\prod_{k=0}^{3}\delta(\omega_k-\Omega_{\sigma(k)}), \tag{B45}$$

and we find

$$M_j(\boldsymbol{\Omega}) = \frac{1}{16\sqrt{4!}}\sum_{\sigma\in S_4}\prod_{k=0}^{3}\phi^*_{\sigma(k)}(\Omega_k)\exp(i\Omega_k\tau_{\sigma(k)}), \tag{B46}$$

where S_4 is the symmetric group on the four letters, $0,\ldots,3$. Therefore, the POVM operator $\hat{E}_{4000}$ is given by

TABLE V. Generated GHZ states depending on detection pattern.

Detection pattern	Generated state
0, 1	$\frac{1}{\sqrt{2}}(\|1010\rangle - \|0101\rangle)$
0, 2	$\frac{1}{\sqrt{2}}(\|1100\rangle - \|0011\rangle)$
0, 3	$\frac{1}{\sqrt{2}}(\|1001\rangle - \|0110\rangle)$
1, 2	$\frac{1}{\sqrt{2}}(-\|1001\rangle + \|0110\rangle)$
1, 3	$\frac{1}{\sqrt{2}}(-\|1100\rangle + \|0011\rangle)$
2, 3	$\frac{1}{\sqrt{2}}(-\|1010\rangle + \|0101\rangle)$

$$\hat{E}_{4000} = \frac{1}{6144}\sum_{\sigma,\sigma' \in S_4}\prod_{k=0}^{3}\mu_{\sigma(k)\sigma'(k)}|0123\rangle_{\text{in}}\langle 0123|. \quad \text{(B47)}$$

3. POVM for GHZ states

The four-partite GHZ state is generated by the same interferometer used for the W-state generation, but two-photon detection is required [78]. Table V lists the detector indices corresponding to the measurement outcomes and generated GHZ states. In this appendix, we consider the first row in Table V as an example. The corresponding projectors at the output side, including multiphoton detection, are as follows:

$$\hat{\mathcal{P}}_{1100} = \hat{\mathcal{P}}_0^{(1)} \otimes \hat{\mathcal{P}}_1^{(1)}, \quad \text{(B48)}$$

$$\hat{\mathcal{P}}_{1200} = \hat{\mathcal{P}}_0^{(1)} \otimes \hat{\mathcal{P}}_1^{(2)}, \quad \text{(B49)}$$

$$\hat{\mathcal{P}}_{2100} = \hat{\mathcal{P}}_0^{(2)} \otimes \hat{\mathcal{P}}_1^{(1)}, \quad \text{(B50)}$$

$$\hat{\mathcal{P}}_{2200} = \hat{\mathcal{P}}_0^{(2)} \otimes \hat{\mathcal{P}}_1^{(2)}, \quad \text{(B51)}$$

$$\hat{\mathcal{P}}_{1300} = \hat{\mathcal{P}}_0^{(1)} \otimes \hat{\mathcal{P}}_1^{(3)}, \quad \text{(B52)}$$

$$\hat{\mathcal{P}}_{3100} = \hat{\mathcal{P}}_0^{(3)} \otimes \hat{\mathcal{P}}_1^{(1)}, \quad \text{(B53)}$$

where

$$\hat{\mathcal{P}}_k^{(1)} = \int d\Omega_k\, \hat{r}_k^\dagger(\Omega_k)|0\rangle_k\langle 0|\hat{r}_k(\Omega_k), \quad \text{(B54)}$$

$$\hat{\mathcal{P}}_k^{(2)} = \frac{1}{2}\iint d\Omega_{k0}d\Omega_{k1}\, \hat{r}_k^\dagger(\Omega_{k0})\hat{r}_k^\dagger(\Omega_{k1})|0\rangle_k\langle 0|\hat{r}_k(\Omega_{k0})\hat{r}_k(\Omega_{k1}), \quad \text{(B55)}$$

$$\hat{\mathcal{P}}_k^{(3)} = \frac{1}{6}\int d^3\boldsymbol{\Omega}_k \bigotimes_{l=0}^{2}\hat{r}_k^\dagger(\Omega_{kl})|0\rangle_k\langle 0|\hat{r}_k(\Omega_{kl}). \quad \text{(B56)}$$

We derive the POVM operators E_{1100}, E_{1200}, E_{2100}, E_{2200}, E_{1300}, and E_{3100} in the rest of this subsection.

a. $\hat{E}_{1100}$

We derive the POVM $\hat{E}_{1100}$ corresponding to the projector $\hat{\mathcal{P}}_{1100}$. In Eq. (B22), the corresponding measurement outcomes are $n_0 = n_1 = 1$ and $n_2 = n_3 = 0$, while the input states are given by Eq. (B34). The elements $M_{\boldsymbol{j}1100}$ of the POVM $\hat{E}_{1100}$ are given by

$$\begin{aligned} &M_{\boldsymbol{j}1100}(\Omega_0,\Omega_1) \\ &= {}_{\text{out}}\langle 0|\left(\sum_{k=0}^{1}V_{j_0k}\hat{R}_{j_0k}\right)\left(\sum_{k=0}^{1}V_{j_1k}\hat{R}_{j_1k}\right) \\ &\quad\times \hat{r}_0^\dagger(\Omega_0)\hat{r}_1^\dagger(\Omega_1)|0\rangle_{\text{out}}. \end{aligned} \quad \text{(B57)}$$

The elements of the beam-splitter matrix $\hat{V}$ (B2) satisfy the following relations:

$$V_{j0} = \frac{1}{2}, \qquad V_{j1} = \frac{(-1)^j}{2}, \quad \text{(B58)}$$

which leads to

$$\begin{aligned} &M_{\boldsymbol{j}1100}(\Omega_0,\Omega_1) \\ &= \frac{1}{4}{}_{\text{out}}\langle 0|[(-1)^{j_1}\hat{R}_{j_0 0}\hat{R}_{j_1 1} + (-1)^{j_0}\hat{R}_{j_0 1}\hat{R}_{j_1 0}] \\ &\quad\times \hat{r}_0^\dagger(\Omega_0)\hat{r}_1^\dagger(\Omega_1)|0\rangle_{\text{out}} \\ &= \frac{1}{4}{}_{\text{out}}\langle 0|\sum_{\sigma\in S_2}(-1)^{j_{\sigma(1)}}\hat{R}_{j_{\sigma(0)}0}\hat{R}_{j_{\sigma(1)}1} \\ &\quad\times \hat{r}_0^\dagger(\Omega_0)\hat{r}_1^\dagger(\Omega_1)|0\rangle_{\text{out}}. \end{aligned} \quad \text{(B59)}$$

Using the commutation relation in Eq. (B7),

$$\begin{aligned} &{}_{\text{out}}\langle 0|\hat{R}_{j_{\sigma(0)}0}\hat{r}_0^\dagger(\Omega_0)|0\rangle_{\text{out}} \\ &\quad = \phi^*_{j_{\sigma(0)}}(\Omega_0)\exp[i\Omega_0\tau_{j_{\sigma(0)}}], \end{aligned} \quad \text{(B60)}$$

we obtain

$$\begin{aligned} &M_{\boldsymbol{j}1100}(\Omega_0,\Omega_1) \\ &= \frac{1}{4}\sum_{\sigma\in S_2}(-1)^{j_{\sigma(1)}}\prod_{k=0}^{1}\phi^*_{j_{\sigma(k)}}(\Omega_k)\exp[i\Omega_k\tau_{j_{\sigma(k)}}]. \end{aligned} \quad \text{(B61)}$$

Substituting Eq. (B61) into (B18), we find

$$\begin{aligned} \hat{E}_{1100} = &\frac{1}{16}\sum_{\boldsymbol{j}\boldsymbol{j}'}|\boldsymbol{j}\rangle_{\text{in}}\langle\boldsymbol{j}'| \\ &\times\left[\sum_{\sigma,\sigma'\in S_2}(-1)^{j_{\sigma(1)}+j'_{\sigma'(1)}}\prod_{k=0}^{1}\mu_{j_{\sigma(k)}j'_{\sigma'(k)}}\right], \end{aligned} \quad \text{(B62)}$$

where $|\boldsymbol{j}\rangle$ and $|\boldsymbol{j}'\rangle$ belong to the same set given in Eq. (B34).

b. $\hat{E}_{1200}$ and $\hat{E}_{2100}$

The POVM $\hat{E}_{1200}$ corresponds to the measurement outcomes $n_0 = 1$, $n_1 = 2$, and $n_2 = n_3 = 0$. In this case, the input states are given by Eq. (B39). The elements M_{j1200} of POVM $\hat{E}_{1200}$ are given by

$$M_{j1200}(\boldsymbol{\Omega}) = \frac{1}{\sqrt{2}}{}_{\text{out}}\langle 0| \bigotimes_{i=0}^{2}\left(\sum_{k=0}^{1} V_{j_i k}\hat{R}_{j_i k}\right) \times \hat{r}_0^{\dagger}(\Omega_{00})\hat{r}_1^{\dagger}(\Omega_{10})\hat{r}_1^{\dagger}(\Omega_{11})|0\rangle_{\text{out}}. \tag{B63}$$

Using Eq. (B58), this reduces to

$$\begin{aligned} M_{j1200}(\boldsymbol{\Omega}) &= \frac{1}{8\sqrt{2}_{\text{out}}}\langle 0|[(-1)^{j_1+j_2}\hat{R}_{j_0 0}\hat{R}_{j_1 1}\hat{R}_{j_2 1} \\ &\quad + (-1)^{j_0+j_1}\hat{R}_{j_0 1}\hat{R}_{j_1 1}\hat{R}_{j_2 0} \\ &\quad +(-1)^{j_0+j_2}\hat{R}_{j_0 1}\hat{R}_{j_1 0}\hat{R}_{j_2 1}] \\ &\quad \times \hat{r}_0^{\dagger}(\Omega_{00})\hat{r}_1^{\dagger}(\Omega_{10})\hat{r}_1^{\dagger}(\Omega_{11})|0\rangle_{\text{out}}. \end{aligned} \tag{B64}$$

Applying the commutation relations in Eqs. (B7) and (B36), we obtain

$$M_{j1200}(\boldsymbol{\Omega}) = \frac{1}{8\sqrt{2}}\sum_{\sigma\in S_3}(-1)^{j_{\sigma(1)}+j_{\sigma(2)}} \times \prod_{k=0}^{2}\phi^{*}_{j_{\sigma(k)}}(\Omega_k)\exp i\Omega_k\tau_{j_{\sigma(k)}}. \tag{B65}$$

Therefore, the POVM $\hat{E}_{1200}$ is given by

$$\hat{E}_{1200} = \frac{1}{128}\sum_{\boldsymbol{j}\boldsymbol{j}'}|\boldsymbol{j}\rangle_{\text{in}}\langle\boldsymbol{j}'| \left[\sum_{\sigma,\sigma'\in S_3}(-1)^{j_{\sigma(1)}+j_{\sigma(2)}+j'_{\sigma'(1)}+j'_{\sigma'(2)}}\prod_{k=0}^{2}\mu_{j_{\sigma(k)}j'_{\sigma'(k)}}\right]. \tag{B66}$$

where $|\boldsymbol{j}\rangle$ and $|\boldsymbol{j}'\rangle$ belong to the same set in Eq. (B39).

For the POVM $\hat{E}_{2100}$, indices of output ports are different from those in $\hat{E}_{1200}$, while the number of detected photons is the same. Thus, the structure of $\hat{E}_{2100}$ is identical to that of $\hat{E}_{1200}$ except for the phase factor. Since only one photon is detected in the first port, two of the four (-1)-exponent terms in Eq. (B66) are removed, resulting

$$\hat{E}_{2100} = \frac{1}{128}\sum_{\boldsymbol{j}\boldsymbol{j}'}|\boldsymbol{j}\rangle_{\text{in}}\langle\boldsymbol{j}'| \times\left[\sum_{\sigma,\sigma'\in S_3}(-1)^{j_{\sigma(2)}+j'_{\sigma'(2)}}\prod_{k=0}^{2}\mu_{j_{\sigma(k)}j'_{\sigma'(k)}}\right]. \tag{B67}$$

c. $\hat{E}_{2200}$

The POVM $\hat{E}_{2200}$ corresponds to the measurement outcomes $n_0 = n_1 = 2$ and $n_2 = n_3 = 0$. Since photons are injected into all input ports, the input state is given by $|\boldsymbol{j}\rangle = |1234\rangle$. The elements M_{j2200} of POVM $\hat{E}_{2200}$ are given by

$$M_{j2200}(\boldsymbol{\Omega}) = \frac{1}{2}{}_{\text{out}}\langle 0| \bigotimes_{i=0}^{3}\left(\sum_{k=0}^{1} V_{j_i k}\hat{R}_{j_i k}\right) \times \hat{r}_0^{\dagger}(\Omega_{00})\hat{r}_0^{\dagger}(\Omega_{01})\hat{r}_1^{\dagger}(\Omega_{10})\hat{r}_1^{\dagger}(\Omega_{11})|0\rangle_{\text{out}}. \tag{B68}$$

Using Eq. (B58) and the commutation relation in Eq. (B36), this reduces to

$$M_{j2200}(\boldsymbol{\Omega}) = \frac{1}{2^5}\sum_{\sigma\in S_4}(-1)^{\sigma(2)+\sigma(3)} \times \prod_{k=0}^{3}\phi^{*}_{\sigma(k)}(\Omega_k)\exp i\Omega_k\tau_{\sigma(k)}, \tag{B69}$$

where the subscripts are replaced with $0, \ldots, 3$ for simplicity. Therefore, the POVM $\hat{E}_{2200}$ is given by

$$\hat{E}_{2200} = \frac{1}{2^{10}}|1234\rangle_{\text{in}}\langle 1234| \times \sum_{\sigma,\sigma'\in S_4}(-1)^{\sigma(2)+\sigma(3)+\sigma'(2)+\sigma'(3)}\prod_{k=0}^{3}\mu_{\sigma(k),\sigma'(k)}. \tag{B70}$$

d. $\hat{E}_{1300}$ and $\hat{E}_{3100}$

The POVM $\hat{E}_{1300}$ corresponds to the measurement outcomes $n_0 = 1$, $n_1 = 3$, and $n_2 = n_3 = 0$. Since photons are injected into all input ports, the input state is given by $|\boldsymbol{j}\rangle = |1234\rangle$. The elements M_{j1300} of POVM $\hat{E}_{1300}$ are expressed as

$$M_{j1300}(\boldsymbol{\Omega}) = \frac{1}{\sqrt{6}}{}_{\text{out}}\langle 0| \bigotimes_{i=0}^{3}\left(\sum_{k=0}^{1} V_{j_i k}\hat{R}_{j_i k}\right) \times \hat{r}_0^{\dagger}(\Omega_{00})\hat{r}_1^{\dagger}(\Omega_{10})\hat{r}_1^{\dagger}(\Omega_{11})\hat{r}_1^{\dagger}(\Omega_{12})|0\rangle_{\text{out}}. \tag{B71}$$

Using Eq. (B58) and the commutation relation in Eq. (B41), we obtain

$$M_{j1300}(\boldsymbol{\Omega}) = \frac{1}{16\sqrt{6}}\sum_{\sigma\in S_4}(-1)^{\sigma(1)+\sigma(2)+\sigma(3)} \times \prod_{k=0}^{3}\phi^{*}_{\sigma(k)}(\Omega_k)\exp i\Omega_k\tau_{\sigma(k)}. \tag{B72}$$

where the subscripts are replaced with $0,\ldots,3$ for simplicity. The POVM $\hat{E}_{1300}$ becomes

$$\hat{E}_{1300} = \frac{1}{2^8\cdot 6}|1234\rangle_{\text{in}}\langle 1234| \times \sum_{\sigma,\sigma'\in S_4}(-1)^{\sum_{i=1}^{3}\sigma(i)+\sigma'(i)}\prod_{k=0}^{3}\mu_{\sigma(k)\sigma'(k)}. \quad \text{(B73)}$$

As in the derivation of $\hat{E}_{2100}$, the POVM $\hat{E}_{3100}$ has the same structure as $\hat{E}_{1300}$ except for the phase factors. Since only one photon is measured in the first output port for $\hat{E}_{3100}$, four of the six (-1)-exponent terms in Eq. (B73) are removed, leading to

$$\hat{E}_{3100} = \frac{1}{2^8\cdot 6}|1234\rangle_{\text{in}}\langle 1234| \times \sum_{\sigma,\sigma'\in S_4}(-1)^{\sigma(3)+\sigma'(3)}\prod_{k=0}^{3}\mu_{\sigma(k)\sigma'(k)}. \quad \text{(B74)}$$

4. POVM for Bell states

In order to create a Bell state on two emitters, either A-B or C-D, we adopt a set of POVMs $\{\hat{E}_{ij}\}$ that have been previously derived in the literature and are directly applicable to our construction [108]. Explicitly, the elements of the POVM are given as follows (assuming pairwise indistinguishability as μ):

$$\hat{E}_{00} = \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \quad \text{(B75)}$$

$$\hat{E}_{10} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 1 & \mu & 0\\ 0 & \mu^* & 1 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \quad \text{(B76)}$$

$$\hat{E}_{01} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 1 & -\mu & 0\\ 0 & -\mu^* & 1 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \quad \text{(B77)}$$

$$\hat{E}_{11} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1-|\mu|^2\end{pmatrix}, \quad \text{(B78)}$$

$$\hat{E}_{20} = \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1+|\mu|^2\end{pmatrix}, \quad \text{(B79)}$$

$$\hat{E}_{02} = \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1+|\mu|^2\end{pmatrix}. \quad \text{(B80)}$$

The above operators satisfy the completeness relation

$$\sum_{i,j}E_{ij} = \mathbb{I},$$

ensuring a valid POVM. When we create an elementary Bell state on the emitters, these POVMs are employed for the detectors, pairwise, either for the PNR or the non-PNR case, as described in Appendix A2f.

APPENDIX C: MAPPING HARDWARE NOISE TO QEC PERFORMANCE

1. Superoperator and Pauli twirling

Let $\mathcal{S}$ denote the noisy quantum map that implements a stabilizer measurement cycle on the data qubits' subsystem $\mathcal{H}_D$ (including the interaction with communication qubits and their projective readout), producing a (measurement outcome-conditioned) reduced state on the data qubits. The key idea is that $\mathcal{S}$ has all the noise due to the hardware, and then we want to map its action to a distribution of pure Pauli errors on the data qubits that can be applied to a surface code again. The Choi-Jamiolkowski isomorphism accomplishes this action as it sets up a relationship between the state and the noisy quantum channel. The Choi matrix [121,122] of $\mathcal{S}$ is

$$J_{\mathcal{S}} = (\mathcal{S}\otimes\mathbb{I})(|\Phi\rangle\langle\Phi|), \qquad |\Phi\rangle = \frac{1}{\sqrt{2^{n_D}}}\sum_{i=0}^{2^{n_D}-1}|i\rangle\otimes|i\rangle, \quad \text{(C1)}$$

where n_D is the number of data qubits in the stabilizer. Here, $|i\rangle|i\rangle$ denotes joint qubit computational basis states over the stabilizer unit-cell data qubits and the auxiliary system, which together form Bell pairs, as described above, and can be expanded and also be written in the binary form. For instance, if $n_D = 4$, then $\rho = \frac{1}{4}|\Phi_+\rangle^{\otimes 4} = \frac{1}{4}(|0000\rangle|0000\rangle + \cdots + |1111\rangle|1111\rangle)$. The map action can be recovered from $J_{\mathcal{S}}$ via

$$\mathcal{S}(\rho) = \text{Tr}_A[(\mathbb{I}\otimes\rho^T)J_{\mathcal{S}}], \quad \text{(C2)}$$

or equivalently in vectorized form $\text{vec}[\mathcal{S}(\rho)] = \mathcal{S}_{\text{super}}\text{vec}[\rho]$ with $\mathcal{S}_{\text{super}}$ obtained from $J_{\mathcal{S}}$. Here $\text{Tr}_A[\cdot]$ is the partial trace over subsystem A.

A convenient operator-sum (Kraus) representation is obtained by diagonalizing $J_{\mathcal{S}}$,

$$J_{\mathcal{S}} = \sum_j \lambda_j |v_j\rangle\langle v_j|, \qquad K_j = \sqrt{\lambda_j}\,\mathrm{mat}(|v_j\rangle), \tag{C3}$$

where $\mathrm{mat}(\cdot)$ reshapes the Choi eigenvector into an operator on $\mathcal{H}_D$. The channel is then $\mathcal{S}(\rho) = \sum_j K_j \rho K_j^\dagger$. Which means that we can see the action of $\mathcal{S}$ in terms of its decomposition into Kraus operators. We can either find this decomposition or select a valid choice that we precompute.

Since $\mathcal{S}$ includes distinct classical outcomes (e.g., measurement result $m = \pm 1$, GHZ success/failure flag g), we build separate conditioned Choi matrices $J_{\mathcal{S}}^{(g,m)}$ and repeat the procedure for each (g, m) branch; the final tabulation is organized per branch in separate columns of various possible combinations with their weights denoting the probabilities.

Pauli twirling maps a general CPTP (completely positive and trace preserving) map into a Pauli channel by averaging conjugation over the n-qubit Pauli group $\mathcal{P}_n$ [123,124]. Concretely, the twirled channel is

$$\mathcal{S}_{\mathrm{PT}}(\rho) = \frac{1}{4^{n_D}} \sum_{P \in \mathcal{P}_{n_D}} P\,\mathcal{S}(P\rho P)P. \tag{C4}$$

The twirl eliminates off-diagonal operator terms in a Pauli operator expansion and yields a Pauli channel of the form

$$\mathcal{S}_{\mathrm{PT}}(\rho) = \sum_e q_e\, E_e\, \rho\, E_e^\dagger, \qquad E_e \in \mathcal{P}_{n_D}, \tag{C5}$$

with non-negative probabilities q_e satisfying $\sum_e q_e = 1$ (for trace-preserving conditioned maps, the normalization may be a subunit and interpreted as conditional probability for configurations in g, m variables).

A numerically robust route to compute $\{q_e\}$ is the following (used in our code for noisy circuit simulation to superoperator decomposition):

1. Diagonalize the Choi matrix and obtain Kraus operators $\{K_j\}$ as above (work with each conditioned Choi $J_{\mathcal{S}}^{(g,m)}$ separately).
2. Expand each Kraus operator in the Pauli basis $\{E_e\}$, using the Hilbert-Schmidt inner product. For w data qubits, the Pauli strings obey $\mathrm{Tr}(E_e^\dagger E_{e'}) = 2^w \delta_{ee'}$. The expansion coefficients are

$$c_{j,e} = \frac{1}{2^w}\mathrm{Tr}(E_e^\dagger K_j), \qquad K_j = \sum_e c_{j,e} E_e.$$

3. The Pauli-twirled channel retains only diagonal terms in this basis, and the Pauli error weights are

$$q_e = \sum_j |c_{j,e}|^2. \tag{C6}$$

   Equivalently, expanding the channel $\mathcal{S}(\rho) = \sum_{e,e'} \alpha_{e,e'} E_e \rho E_{e'}^\dagger$, the twirl sets $\alpha_{e,e'} \mapsto 0$ for $e \neq e'$ and $q_e = \alpha_{e,e}$.
4. Enforce numerical trace preservation or normalization if needed (due to numerical errors and precision): for conditioned maps (e.g., conditioned on an observed measurement outcome) the q_e sum to the conditional trace $\mathrm{Tr}[\mathcal{S}(\mu/2^w)]$; we divide by this trace if a normalized conditional probability vector is required.

The procedure above is stable and avoids explicit analytic evaluation. This procedure is repeated for many iterations for both X- and Z-type distributed stabilizer circuits. It produces the operationally meaningful probabilities (or error configuration weights) $\{q_e\}$ that approximate the original noisy map (directly mapped to hardware parameters) by the Pauli channel that is closest under the Pauli-twirl ansatz.

The twirled decomposition is directly used to build the compact table consumed by the surface code QEC simulator in the next step. For each conditioned branch (GHZ success/failure g, measurement outcome m), we list all Pauli strings E_e acting on the w data qubits together with their probability weight $q_e^{(g,m)}$. The effective (noisy) stabilizer projector then reads

$$\begin{aligned}\Pi_m^{\mathrm{eff}}(\rho_D) = &\sum_e q_e^{(\mathrm{succ},m)}\, E_e\, \Pi_m^{\mathrm{ideal}}(\rho_D) E_e^\dagger \\ &+ \sum_e q_e^{(\mathrm{fail},m)}\, E_e\, \Pi_m^{\mathrm{ideal}}(\rho_D) E_e^\dagger \\ &+ \sum_e q_e^{(\mathrm{succ},\bar{m})}\, E_e\, \Pi_{\bar{m}}^{\mathrm{ideal}}(\rho_D) E_e^\dagger \\ &+ \sum_e q_e^{(\mathrm{fail},\bar{m})}\, E_e\, \Pi_{\bar{m}}^{\mathrm{ideal}}(\rho_D) E_e^\dagger,\end{aligned} \tag{C7}$$

where $\bar{m}$ denotes the complementary measurement outcome when the stabilizer measurement circuit fails. The first (second) two summations correspond to the branch producing outcomes with GHZ success and failure, respectively, for m (and outcome with measurement error $\bar{m}$ for error configuration E_e). All these numerical computations are achieved by our source code CIRCUITSIMULATOR tailored for the single-shot emission scheme. The output is the table of format superoperators that can be used as explained in the next subsection.

2. Monte Carlo simulations over superoperator for QEC performance

This section describes the numerical procedure used to estimate logical success and error rates and fault-tolerance thresholds by Monte Carlo sampling of the precomputed tabular superoperator and subsequent decoding. The recipe below is the exact implementation used in our QSURFACE workflow.

The tabular superoperator supplies a discrete probability distribution over triples (E_e, g, m) where $E_e \in \mathcal{P}_w$ is a Pauli string on the w data qubits of a stabilizer unit cell, $g \in \{\text{succ}, \text{fail}\}$ is the GHZ flag (or other branch flag), and $m \in \{\pm 1\}$ is the reported measurement outcome (possibly flipped by measurement error). The table provides probabilities $q_e^{(g,m)}$ for each triple. These tables are provided in the source code and accompanying data as CSV files. For a visual representation of these superoperator tables, one can also refer to Ref. [6]. A single Monte Carlo shot consists of sampling one row from the table for every stabilizer measurement event occurring in a QEC subround (of either X or Z type), applying the sampled errors to the code data qubits, assembling the syndrome record, and decoding.

Spatial layout and scheduling: Given a code of physical distance d, we construct the lattice of data qubits and stabilizer unit cells for one QEC round. For distributed measurement circuits, we expand each stabilizer into the sequence of subrounds (sublayers) required by the hardware scheduling (see main text). Each stabilizer measurement attempt corresponds to one draw from the table.

Sampling procedure: For each stabilizer measurement event (plaquette-Z or vertex-X, at a particular time or subround), we:

1. Select the relevant conditioned table: choose the rows corresponding to the stabilizer geometry (which fixes w qubits' ordering) and the measurement instrument branches (GHZ success/fail).
2. Draw a random sample from the discrete distribution $\{q_e^{(g,m)}\}$. Practically, we accumulate the cumulative distribution and perform a uniform random number generation $u \in [0, 1)$ to select the row where u falls.
3. Interpret the sampled triple (E_e, g, m): apply the Pauli string E_e to the specified data qubits (composition is multiplicative with prior errors), record the measurement outcome m as the stabilizer result for this subround, and record the GHZ success flag g.

GHZ failure handling and cutoff policy: If $g = \text{fail}$ for a stabilizer subround, the code employs the decision rule for aborting that stabilizer measurement: the stabilizer measurement is aborted and the previous valid outcome is reused (or a default outcome is substituted) depending on the chosen protocol. This logic is implemented exactly during sampling by maintaining a per-stabilizer memory of the last accepted outcome and updating it only on successful GHZ events or according to the protocol's replacement rule.

Order of application: When multiple sampled Pauli strings act on overlapping data qubits (from different stabilizer attempts within a QEC cycle), we apply them in the same chronological order as the physical protocol: for each subround, we (i) sample and apply data-qubit Pauli from that stabilizer, (ii) update the stored stabilizer outcome, then advance to the next stabilizer. This preserves time ordering and correctly models idling noise via the decoherence scheduling described in Appendix A.

Syndrome assembly: After all subrounds of a QEC cycle are sampled and applied, the simulation produces a set of stabilizer outcomes per round (including reused or substituted outcomes where GHZ failed). These outcomes form the space–time syndrome record (a 3D syndrome graph over subrounds in time layers).

Decoder input: We pass the syndrome record to the chosen decoder (Union-Find [87,125,126] or minimum weight perfect matching [127,128]). The decoder inputs are simply binary stabilizer outcomes per stabilizer location and time; it does not require knowledge of the underlying physical error mechanisms because the superoperator sampling already embeds those mechanisms into the syndrome. Note that we use a weighted UF decoder version for the thresholds reported in this work.

Logical outcome and success metric: The decoder returns a correction operator C. We apply C (classically) to the simulated data-qubit error configuration and check whether the total error combined with C implements a logical operator on the code (i.e., whether the net logical Pauli is nontrivial). A Monte Carlo shot is declared a logical failure if a logical error occurs. Otherwise, it is a logical success. Formally:

$$\text{logical failure} \Longleftrightarrow L_{\text{net}} \neq I,$$

where L_{net} is the equivalence class of the net data-qubit error after applying the decoder correction.

Estimating logical error and success rates (repetition and statistics): For each tuple of hardware parameters and physical error rate p, run N_{shots} independent Monte Carlo trials (typical $N_{\text{shots}} = 5 \times 10^4$). The empirical logical error rate is

$$p_{\text{L}} = \frac{1}{N_{\text{shots}}} \sum_{i=1}^{N_{\text{shots}}} \mathbf{1}\{\text{failure}_i\},$$

and the success rate is $1 - p_{\text{L}}$. The standard error is $\sqrt{p_{\text{L}}(1 - p_{\text{L}})/N_{\text{shots}}}$; choose N_{shots} to ensure desired confidence per data point.

Confidence and rare events: When logical error probabilities are very small, increase N_{shots} or use importance sampling and splitting methods. For our reported thresholds, we used $N_{\text{shots}} = 10^4$, which yields subpercent statistical uncertainty in the threshold estimates reported.

Threshold extraction recipe: To obtain a threshold p_{th} for a given protocol and hardware set:

1. Fix other parameters (coherence times, α, or hardware parameter set ES-x).
2. For a set of physical error rates p spanning a region around the expected threshold, simulate logical

error rates $p_L(p,d)$ for several code distances d (we use $d \in \{4,6,8,10,12\}$ as it makes the distributed toric surface code simulations easier, and the planar surface code performance is expected to be similar).
3. For each d plot $p_L(p,d)$ vs p. The threshold is the critical p where logical error curves cross or become approximately distance independent. Numerically find p_{th} by locating the intersection of $p_L(p,d)$ curves (or by fitting finite-size scaling forms to the data and extracting the critical point).
4. Estimate uncertainty by bootstrap resampling of Monte Carlo trials or by repeating the entire sweep with different random number generator seeds and reporting the standard deviation of the extracted p_{th}.

The method for optimizing for cutoff times and fitting procedure for threshold estimation for the logical success rate curve is explained in Appendix F. The recipe in this appendix allows us to link the hardware parameters directly to QEC performance.

APPENDIX D: SWEEPING BRIGHT-STATE PARAMETER FOR BASE AND DISTILLING STATE

For the memory distillation protocols discussed in the main text, we initially assumed that the bright-state parameter α was identical for both the base state and the distilling state. In general, this assumption is not required, and it is worthwhile to investigate potential trade-offs when α is varied independently for the two states. Intuitively, one might expect that choosing lower values of α could maximize fidelity. However, the impact of the distillation quantum circuit on both the final success probability and fidelity may reveal nontrivial trade-offs.

Figure 13 presents the results for the four memory distillation protocols. Here, we fix the coherence time to $T = 10^6$, employ the ES-2 hardware parameter set, consider a gate error of $p = 10^{-3}$, and assume PNR detectors. The fault-tolerance regions are indicated by the black line-shaded areas. While the output infidelity for this set of input parameters does not reach fault-tolerance—consistent with Fig. 10 for the ES-2 platform—the plots illustrate the variation of the output infidelity with respect to α_{base} and $\alpha_{distill}$.

If the desired infidelity threshold is attainable, there exists flexibility in the choice of α_{base} and $\alpha_{distill}$. For symmetric protocols, where the base and distilling states are of the same type (e.g., distilling GHZ with GHZ or W with W), it is possible to select a relatively high value of α for one of the states. This flexibility can be leveraged to relax experimental requirements, enabling optimization of success probabilities while still meeting infidelity constraints.

APPENDIX E: HEATMAPS FOR HARDWARE AND COHERENCE PARAMETERS

In a similar spirit to Appendix D, we perform a heatmap analysis over coherence times and hardware parameters to gain deeper insights into the precise fault-tolerance requirements of all protocols. In the main text, for numerical tractability, we preselected the most relevant parameters for the two-dimensional plots. Here, we extend this analysis by presenting three-dimensional heatmaps where both coherence times are varied for all hardware parameter sets. The results are shown in Figs. 14 and 15 in two parts, covering all raw, optical, and memory distillation protocols.

Throughout this analysis, we employ PNR detectors, a physical error rate of $p = 10^{-3}$, and set $\alpha = 0.025$, except for the OD-GHZ protocol, where we use $\alpha = 0.5$ to achieve optimal performance as noted in the main text. The coherence times are swept from modest present-day values through near-term improvements to ambitious, futuristic targets. This sweep also covers hardware sets 1, 2, and 3 considered in Ref. [6], and therefore, the results can be directly compared.

The resulting heatmaps (in Figs. 14 and 15) provide additional insight and a roadmap for attaining fault tolerance with single-shot emission-based hardware. Our findings indicate that the primary challenge lies in meeting the stringent infidelity requirements, while the target generation success rates are largely achievable with realistic hardware parameters.

APPENDIX F: THRESHOLD SIMULATIONS

1. Cutoff time optimization

As outlined in the main text, the cutoff time t_{cut} sets the maximum duration allowed for GHZ-state generation during a stabilizer measurement attempt. Any entanglement generation RUS attempts exceeding this time are aborted, and the corresponding stabilizer measurement is handled according to the protocol's failure policy (e.g., outcome substitution or reuse of the previous measurement).

We define $t_{cut}^{100\%}$ as the average time required to achieve a successful GHZ state across all trials (i.e., corresponding to the full success probability over the allowed attempts). More generally, we introduce $t_{cut}^{x\%}$, the cutoff time that captures the average duration required for success within the fastest $x\%$ of attempts. The value $t_{cut}^{100\%}$ is precomputed in a prior simulation stage (when sweeping and scanning the range of physical error rates p); the shorter $t_{cut}^{x\%}$ values are obtained by truncating the distribution of generation times at the xth percentile.

The search for an optimal cutoff proceeds as follows:
1. Choose an initial percentile x—typically $x \in [95\%, 99\%]$.

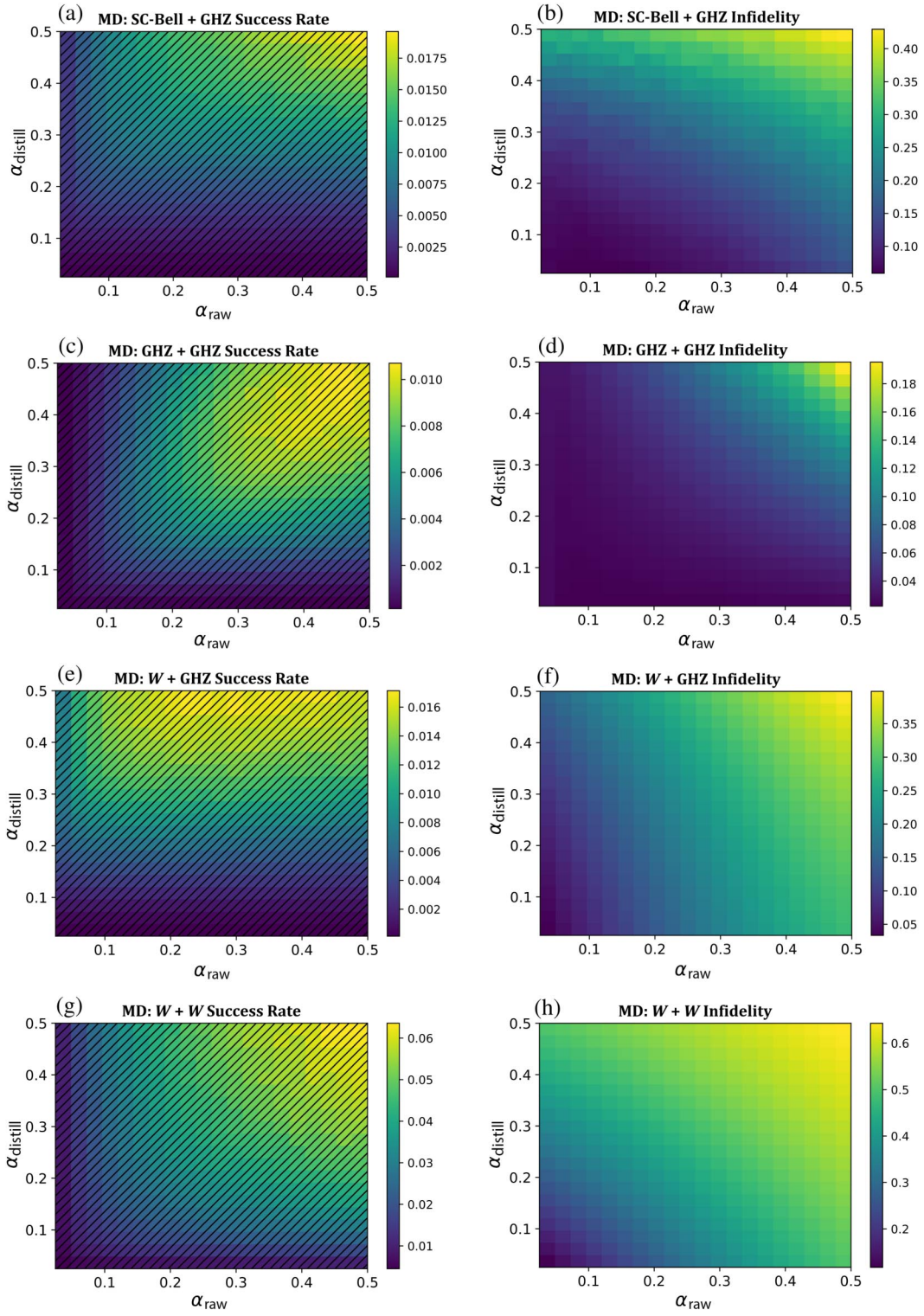


FIG. 13. Bright-state parameter trade-off for protocols with distinct α for base state and distilling state, with PNR detectors. Evaluated with coherence times of $T = 10^6$, hardware parameter set ES-2, and physical errorrates of $p = 10^{-3}$. Fault-tolerance regions are shaded with black lines.

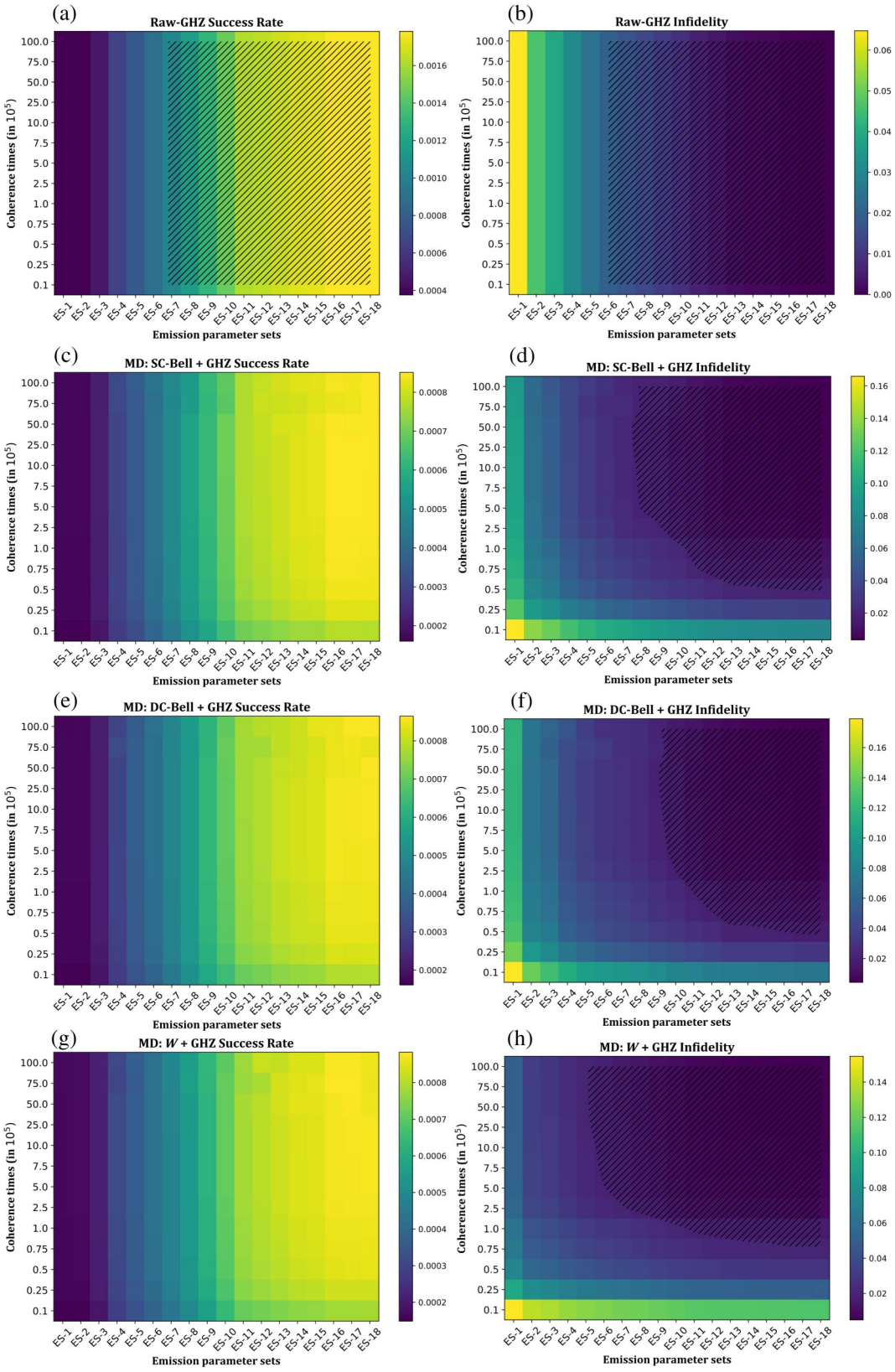


FIG. 14. (1 of 2) Heatmaps for varying coherence times and hardware parameters with PNR detectors, physical error rate of $p = 10^{-3}$, and $\alpha = 0.025$. OD-GHZ is evaluated at $\alpha = 0.5$. Fault-tolerance regions are shaded with black lines.

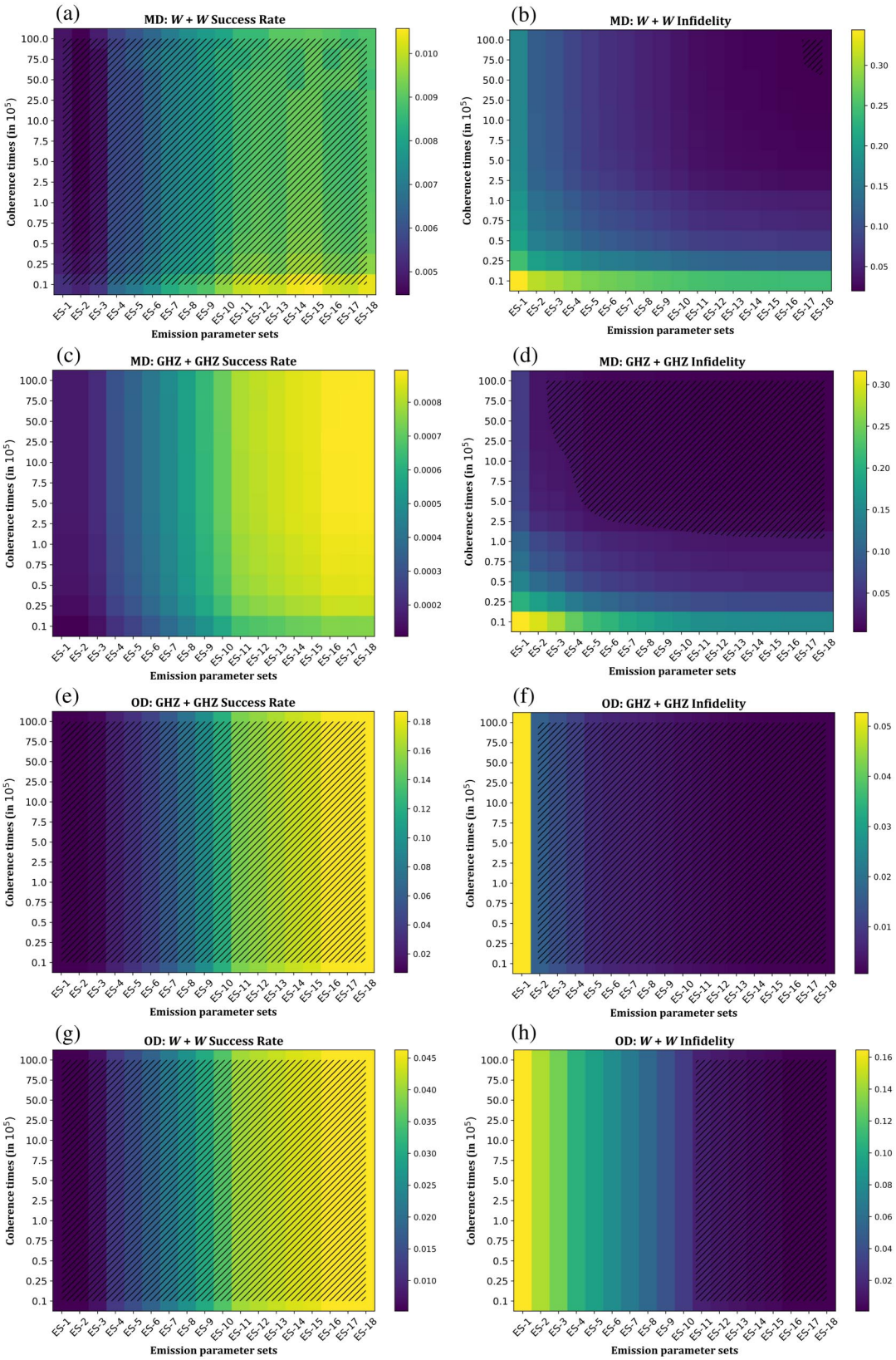


FIG. 15. (2 of 2) Heatmaps for varying coherence times and hardware parameters with PNR detectors, physical error rate of $p = 10^{-3}$, and $\alpha = 0.025$. OD-GHZ is evaluated at $\alpha = 0.5$. Fault-tolerance regions are shaded with black lines.

2. Determine $t_{\text{cut}}^{x\%}$ from the simulated distribution of GHZ-generation times. The CIRCUITSIMULATOR outputs this based on the x input.
3. Run the noisy stabilizer-measurement simulator using this fixed $t_{\text{cut}}^{x\%}$ to estimate the logical error rate p_{L} over a sweep of physical error rates p.
4. If $p_{\text{L}} \lesssim 0.2$ for some p, perform a full threshold search (varying code distance) to estimate p_{th}. If $p_{\text{L}} > 0.80$ for all tested p, discard this x value as the probability of finding a threshold is negligible (verified numerically) with such high logical error rate.

In general, $t_{\text{cut}}^{x\%}$ grows approximately exponentially with x. At low x, increasing x (and thus t_{cut}) improves the threshold p_{th}, as more stabilizer outcomes are successfully collected per QEC cycle. Beyond a certain point, however, p_{th} saturates: additional collection time yields negligible gains in syndrome information but introduces more decoherence noise due to longer waiting times. For very high x, the increased decoherence dominates and p_{th} drops sharply.

We therefore perform a binary search over x to locate the optimal percentile that maximizes p_{th} within statistical error bars. The reported code threshold for each architecture is the highest p_{th} obtained in this search, using the corresponding entanglement-generation scheme. In the next subsection, we highlight the exact fitting procedure that we use to estimate thresholds from the logical success rates for varying code distances.

Figure 16 demonstrates the fundamental trade-off between syndrome information fidelity and memory decoherence in optimizing the GHZ-generation cutoff completion fraction. For both PNR (blue) and non-PNR (orange) detectors, the code threshold p_{th} initially increases with the completion fraction x due to improved stabilizer measurement success, reaching optimal values of $p_{\text{th}}^{\text{PNR}} = 0.2498\%$ and $p_{\text{th}}^{\text{non-PNR}} = 0.1986\%$ at $x = 99.00\%$ and $x = 98.50\%$, respectively (marked by dashed lines). Beyond this point, the threshold degrades sharply as the prolonged waiting times for rare GHZ-generation events accumulate decoherence on memory qubits faster than the marginal improvement in syndrome outcomes can compensate, reflecting the exponential cost of extending $t_{\text{cut}}^{x\%}$ at high percentiles. The approximately 26% superiority of PNR over non-PNR detection underscores the critical importance of photon-number resolution for fault-tolerant performance in distributed emission-based architectures.

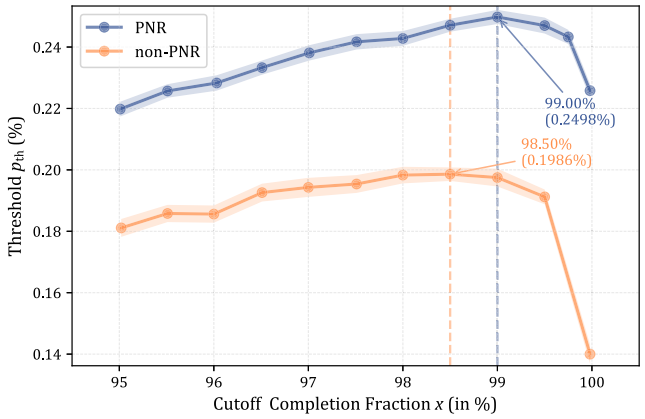


FIG. 16. Code threshold versus cutoff completion fraction for the OD-GHZ protocol under PNR (blue) and non-PNR (orange) detection. Optimal thresholds $p_{\text{th}}^{\text{PNR}} = 0.2498\%$ and $p_{\text{th}}^{\text{non-PNR}} = 0.1986\%$ are attained at $x = 99.00\%$ and $x = 98.50\%$ (marked by dashed lines), respectively. The decline beyond the optimal completion fraction reflects the dominance of decoherence accumulation on idle qubits over gains in syndrome information, establishing the necessity for precise cutoff optimization in distributed surface code.

2. Threshold curve fitting and estimation

In this subsection, we outline the procedure used to determine the threshold values for our quantum error-correcting code simulations, along with their associated uncertainties. The fitting methodology combines Monte Carlo sampling of logical success rates with a weighted nonlinear regression model, followed by statistical error estimation. This numerical recipe is identical to the one followed in Ref. [43].

a. Data acquisition and notation

For a fixed set of physical error parameters, denoted by p, we compute an effective superoperator via the procedure in Appendix C. Using this superoperator, we run Monte Carlo simulations for various surface code distances d to obtain the logical success probability $1 - p_L(p, d) = r_L(p, d)$.

Let each simulation setting be indexed by $i \in \{1, \cdots, n_T\}$, where n_T is the total number of (p, d) combinations simulated. The observed logical success rate is

$$r_i = \frac{M_i}{N_i}, \tag{F1}$$

where N_i is the number of Monte Carlo trials and M_i is the number of trials with no logical failure.

Statistical uncertainty is assumed to follow a binomial distribution, yielding the standard deviation

$$\sigma_i = \sqrt{\frac{r_i(1 - r_i)}{N_i}}. \tag{F2}$$

Uncertainty decreases with increasing N_i, so higher-statistics points naturally carry more weight in the fit.

b. Regression model

Following standard finite-size scaling theory for code thresholds, we fit the logical success rates to Ref. [129]

$$\hat{r}(p,d) = a + b\,(p - p_{\text{th}})d^{1/\kappa} + c\,(p - p_{\text{th}})^2 d^{2/\kappa} + e\,d^{-1/\zeta}, \tag{F3}$$

where $a, b, c, e, p_{\text{th}}, \kappa, \zeta$ are free parameters determined by the fit. Here, p_{th} is the threshold of interest, κ is the finite-size scaling exponent, and ζ governs subleading corrections.

For each data point (p_i, d_i), the residual is

$$\epsilon_i = r_i - \hat{r}(p_i, d_i). \tag{F4}$$

The parameters are found by minimizing the weighted sum of squared residuals

$$Q = \sum_{i=1}^{n_T} \left(\frac{\epsilon_i}{\sigma_i}\right)^2, \tag{F5}$$

which ensures that data with higher statistical precision (smaller σ_i) contribute more strongly to the optimization.

c. Weighted nonlinear least squares

We solve the minimization of Eq. (F5) using the Gauss-Newton algorithm for nonlinear regression [130,131]. Let the parameter vector at iteration t be

$$\boldsymbol{\beta}^{(t)} = \left[a^{(t)} \;\; b^{(t)} \;\; c^{(t)} \;\; d^{(t)} \;\; p_{\text{th}}^{(t)} \;\; \kappa^{(t)} \;\; \zeta^{(t)}\right]^T \tag{F6}$$

and define the parameter update $\Delta\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}$.

Linearizing Eq. (F3) around $\boldsymbol{\beta}^{(t)}$ yields the iterative update rule

$$\Delta\boldsymbol{\beta}^{(t+1)} = [(\mathbf{J}^{(t)})^T \boldsymbol{\Sigma}^{-1} \mathbf{J}^{(t)}]^{-1} (\mathbf{J}^{(t)})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}^{(t)}, \tag{F7}$$

where $\mathbf{J}^{(t)}$ is the Jacobian matrix of partial derivatives $\partial\hat{r}/\partial\beta_j$ evaluated at step t, $\boldsymbol{\epsilon}^{(t)}$ is the residual vector at step t, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \cdots, \sigma_{n_c}^2)$.

The iteration proceeds until $\boldsymbol{\beta}^{(t)}$ converges within a chosen tolerance.

d. Estimating parameter uncertainties

Once convergence is reached, the covariance matrix of the fitted parameters is estimated as

$$\text{Var}(\hat{\boldsymbol{\beta}}) \approx [\mathbf{J}^T \boldsymbol{\Sigma}^{-1} \mathbf{J}]^{-1}, \tag{F8}$$

where $\mathbf{J}$ is the Jacobian matrix at the converged parameter values.

The quality of the fit is assessed using the reduced chi-squared statistic

$$\chi_\nu^2 = \frac{Q}{\nu}, \qquad \nu = n_T - n_P, \tag{F9}$$

where $n_P = 7$ is the number of fitting parameters. A value $\chi_\nu^2 \approx 1$ indicates a statistically consistent fit; $\chi_\nu^2 > 1$ may suggest model inadequacy or underestimated σ_i.

If $\chi_\nu^2 > 1$, we rescale $\boldsymbol{\Sigma}$ by χ_ν^2 to obtain more conservative uncertainty estimates:

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \chi_\nu^2 [\mathbf{J}^T \boldsymbol{\Sigma}^{-1} \mathbf{J}]^{-1}. \tag{F10}$$

3. Confidence intervals for p_{th}

The standard deviation of $\hat{p}_{\text{th}}$ is given by the square root of the corresponding diagonal entry of $\text{Var}(\hat{\boldsymbol{\beta}})$. Confidence intervals are obtained by multiplying this standard deviation by the student's t factor t_{ci} corresponding to the desired confidence level and ν degrees of freedom.

For the 95% confidence intervals shown in our threshold plots, $t_{\text{ci}} \approx 1.96$ for large ν, with slightly larger values for small sample sizes.

[1] C. Gidney and M. Ekerå, How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits, Quantum **5**, 433 (2021).

[2] C. Gidney, How to factor 2048 bit RSA integers with less than a million noisy qubits, arXiv:2505.15917.

[3] M. Mohseni *et al.*, How to build a quantum supercomputer: Scaling from hundreds to millions of qubits, arXiv:2411.10406.

[4] A. Megrant and Y. Chen, Scaling up superconducting quantum computers, Nat. Electron. **8**, 549 (2025).

[5] X. Wu, H. Yan, G. Andersson, A. Anferov, M.-H. Chou, C. R. Conner, J. Grebel, Y. J. Joshi, S. Li, J. M. Miller, R. G. Povey, H. Qiao, and A. N. Cleland, Modular quantum processor with an all-to-all reconfigurable router, Phys. Rev. X **14**, 041030 (2024).

[6] S. Singh, F. Gu, S. de Bone, E. Villaseñor, D. Elkouss, and J. Borregaard, Modular architectures and entanglement schemes for error-corrected distributed quantum computation, npj Quantum Inf. **12**, 3 (2025).

[7] M. Caleffi, M. Amoretti, D. Ferrari, J. Illiano, A. Manzalini, and A. S. Cacciapuoti, Distributed quantum computing: A survey, Comput. Networks **254**, 110672 (2024).

[8] D. Main, P. Drmota, D. P. Nadlinger, E. M. Ainley, A. Agrawal, B. C. Nichol, R. Srinivas, G. Araneda, and D. M. Lucas, Distributed quantum computing across an optical network link, Nature (London) **638**, 383 (2025).

[9] T. J. Yoder, E. Schoute, P. Rall, E. Pritchett, J. M. Gambetta, A. W. Cross, M. Carroll, and M. E. Beverland, Tour de gross: A modular quantum computer based on bivariate bicycle codes, arXiv:2506.03094.

[10] H. Aghaee Rad *et al.*, Scaling and networking a modular photonic quantum computer, Nature (London) **638**, 912 (2025).

[11] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects, Phys. Rev. A **89**, 022317 (2014).
[12] M. Caleffi, A. S. Cacciapuoti, and G. Bianchi, Quantum internet: from communication to distributed computing!, in *Proceedings of the 5th ACM International Conference on Nanoscale Computing and Communication, NANOCOM '18* (Association for Computing Machinery, New York, NY, USA, 2018).
[13] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, Quantum internet: Networking challenges in distributed quantum computing, IEEE Network **34**, 137 (2020).
[14] R. Van Meter and S. J. Devitt, The path to scalable distributed quantum computing, Computer **49**, 31 (2016).
[15] D. Cuomo, M. Caleffi, and A. S. Cacciapuoti, Towards a distributed quantum computing ecosystem, IET Quantum Commun. **1**, 3 (2020).
[16] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, eaam9288 (2018).
[17] S. Sunami, S. Tamiya, R. Inoue, H. Yamasaki, and A. Goban, Scalable networking of neutral-atom qubits: Nanofiber-based approach for multiprocessor fault-tolerant quantum computers, PRX Quantum **6**, 010101 (2025).
[18] J. Roffe, Quantum error correction: An introductory guide, Contemp. Phys. **60**, 226 (2019).
[19] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, England, 2013).
[20] K. Fujii, Quantum computation with topological codes: From qubit to topological fault-tolerance, arXiv:1504.01444.
[21] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).
[22] R. Acharya *et al.*, Quantum error correction below the surface code threshold, Nature (London) **638**, 920 (2025).
[23] N. Lacroix *et al.*, Scaling and logic in the colour code on a superconducting quantum processor, Nature (London) **645**, 614 (2025).
[24] A. Eickbusch *et al.*, Demonstrating dynamic surface codes, arXiv:2412.14360.
[25] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, Nature (London) **627**, 778 (2024).
[26] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, Logical quantum processor based on reconfigurable atom arrays, Nature (London) **626**, 58 (2024).
[27] Q. Xu, J. P. Bonilla Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasić, M. D. Lukin, L. Jiang, and H. Zhou, Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays, Nat. Phys. **20**, 1084 (2024).
[28] B. Grinkemeyer, E. Guardado-Sanchez, I. Dimitrova, D. Shchepanovich, G. E. Mandopoulou, J. Borregaard, V. Vuletić, and M. D. Lukin, Error-detected quantum operations with neutral atoms mediated by an optical cavity, Science **387**, 1301 (2025).
[29] M. Cain, C. Zhao, H. Zhou, N. Meister, J. P. B. Ataides, A. Jaffe, D. Bluvstein, and M. D. Lukin, Correlated decoding of logical algorithms with transversal gates, Phys. Rev. Lett. **133**, 240602 (2024).
[30] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).
[31] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).
[32] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).
[33] R. Acharya *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).
[34] D. Litinski, A game of surface codes: Large-scale quantum computing with lattice surgery, Quantum **3**, 128 (2019).
[35] M. McEwen, D. Bacon, and C. Gidney, Relaxing hardware requirements for surface code circuits using time-dynamics, Quantum **7**, 1172 (2023).
[36] AdeMarti iOlius, J. E. Martinez, P. Fuentes, P. M. Crespo, and J. Garcia-Frias, Performance of surface codes in realistic quantum hardware, Phys. Rev. A **106**, 062428 (2022).
[37] G. Watkins, H. M. Nguyen, K. Watkins, S. Pearce, H.-K. Lau, and A. Paler, A high performance compiler for very large scale surface code computations, Quantum **8**, 1354 (2024).
[38] A. Erhard, H. Poulsen Nautrup, M. Meth, L. Postler, R. Stricker, M. Stadler, V. Negnevitsky, M. Ringbauer, P. Schindler, H. J. Briegel, R. Blatt, N. Friis, and T. Monz, Entangling logical qubits with lattice surgery, Nature (London) **589**, 220 (2021).
[39] N. Delfosse, Hierarchical decoding to reduce hardware requirements for quantum computing, arXiv:2001.11427.
[40] P. Das, C. A. Pattison, S. Manne, D. Carmean, K. Svore, M. Qureshi, and N. Delfosse, A scalable decoder microarchitecture for fault-tolerant quantum computing, arXiv:2001.06598.
[41] N. H. Nickerson, Y. Li, and S. C. Benjamin, Topological quantum computing with a very noisy network and local error rates approaching one percent, Nat. Commun. **4**, 1756 (2013).
[42] N. H. Nickerson, J. F. Fitzsimons, and S. C. Benjamin, Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links, Phys. Rev. X **4**, 041041 (2014).
[43] S. de Bone, P. Möller, C. E. Bradley, T. H. Taminiau, and D. Elkouss, Thresholds for the distributed surface code in the presence of memory decoherence, AVS Quantum Sci. **6**, 033801 (2024).
[44] A. Strikis and L. Berent, Quantum low-density parity-check codes for modular architectures, PRX Quantum **4**, 020321 (2023).

[45] G. de Gliniasty, P. Hilaire, P.-E. Emeriau, S. C. Wein, A. Salavrakos, and S. Mansfield, A spin-optical quantum computing architecture, Quantum **8**, 1423 (2024).
[46] H. Bombin, I. H. Kim, D. Litinski, N. Nickerson, M. Pant, F. Pastawski, S. Roberts, and T. Rudolph, Interleaving: Modular architectures for fault-tolerant photonic quantum computing, arXiv:2103.08612.
[47] H. Bombín, C. Dawson, Y.-H. Liu, N. Nickerson, F. Pastawski, and S. Roberts, Modular decoding: Parallelizable real-time decoding for quantum computers, arXiv:2303.04846.
[48] J. Sinclair, J. Ramette, B. Grinkemeyer, D. Bluvstein, M. D. Lukin, and V. Vuletić, Fault-tolerant optical interconnects for neutral-atom arrays, Phys. Rev. Res. **7**, 013313 (2025).
[49] E. Sutcliffe, B. Jonnadula, C. Le Gall, A. E. Moylett, and C. M. Westoby, Distributed quantum error correction based on hyperbolic floquet codes, in *Proceedings of the 2025 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2025), pp. 649–657.
[50] J. Stack, M. Wang, and F. Mueller, Assessing teleportation of logical qubits in a distributed quantum architecture under error correction, arXiv:2504.05611.
[51] A. Tchebotareva, S. L. N. Hermans, P. C. Humphreys, D. Voigt, P. J. Harmsma, L. K. Cheng, A. L. Verlaan, N. Dijkhuizen, W. de Jong, A. Dréau, and R. Hanson, Entanglement between a diamond spin qubit and a photonic time-bin qubit at telecom wavelength, Phys. Rev. Lett. **123**, 063601 (2019).
[52] M. Pompili, S. L. N. Hermans, S. Baier, H. K. C. Beukers, P. C. Humphreys, R. N. Schouten, R. F. L. Vermeulen, M. J. Tiggelman, L. dos Santos Martins, B. Dirkse, S. Wehner, and R. Hanson, Realization of a multinode quantum network of remote solid-state qubits, Science **372**, 259 (2021).
[53] S. L. N. Hermans, M. Pompili, H. K. C. Beukers, S. Baier, J. Borregaard, and R. Hanson, Qubit teleportation between non-neighbouring nodes in a quantum network, Nature (London) **605**, 663 (2022).
[54] A. Ruskuc, C.-J. Wu, E. Green, S. L. N. Hermans, W. Pajak, J. Choi, and A. Faraon, Multiplexed entanglement of multi-emitter quantum network nodes, Nature (London) **639**, 54 (2025).
[55] R. E. Evans, M. K. Bhaskar, D. D. Sukachev, C. T. Nguyen, A. Sipahigil, M. J. Burek, B. Machielse, G. H. Zhang, A. S. Zibrov, E. Bielejec, H. Park, M. Lončar, and M. D. Lukin, Photon-mediated interactions between quantum emitters in a diamond nanocavity, Science **362**, 662 (2018).
[56] C. M. Knaut, A. Suleymanzade, Y.-C. Wei, D. R. Assumpcao, P.-J. Stas, Y. Q. Huan, B. Machielse, E. N. Knall, M. Sutula, G. Baranes, N. Sinclair, C. De-Eknamkul, D. S. Levonian, M. K. Bhaskar, H. Park, M. Lončar, and M. D. Lukin, Entanglement of nanophotonic quantum memory nodes in a telecom network, Nature (London) **629**, 573 (2024).
[57] D. D. Awschalom, R. Hanson, J. Wrachtrup, and B. B. Zhou, Quantum technologies with optically interfaced solid-state spins, Nat. Photonics **12**, 516 (2018).
[58] J. Hofmann, M. Krug, N. Ortegel, L. Gérard, M. Weber, W. Rosenfeld, and H. Weinfurter, Heralded entanglement between widely separated atoms, Science **337**, 72 (2012).
[59] D. Hucul, I. V. Inlek, G. Vittorini, C. Crocker, S. Debnath, S. M. Clark, and C. Monroe, Modular entanglement of atomic qubits using photons and phonons, Nat. Phys. **11**, 37 (2015).
[60] L. J. Stephenson, D. P. Nadlinger, B. C. Nichol, S. An, P. Drmota, T. G. Ballance, K. Thirumalai, J. F. Goodwin, D. M. Lucas, and C. J. Ballance, High-rate, high-fidelity entanglement of qubits across an elementary quantum network, Phys. Rev. Lett. **124**, 110501 (2020).
[61] T. van Leent, M. Bock, F. Fertig, R. Garthoff, S. Eppelt, Y. Zhou, P. Malik, M. Seubert, T. Bauer, W. Rosenfeld, W. Zhang, C. Becher, and H. Weinfurter, Entangling single atoms over 33 km telecom fibre, Nature (London) **607**, 69 (2022).
[62] S. Saha, M. Shalaev, J. O'Reilly, I. Goetting, G. Toh, A. Kalakuntla, Y. Yu, and C. Monroe, High-fidelity remote entanglement of trapped atoms mediated by time-bin photons, Nat. Commun. **16**, 2533 (2025).
[63] Y. Li and J. D. Thompson, High-rate and high-fidelity modular interconnects between neutral atom quantum processors, PRX Quantum **5**, 020363 (2024).
[64] S. Kikura, K. Tanji, A. Goban, and S. Sunami, Passive quantum interconnects: High-fidelity quantum networking at higher rates with less overhead, arXiv:2507.01229.
[65] F. Afzal *et al.*, Distributed quantum computing in silicon, arXiv:2406.01704.
[66] C. E. Bradley, S. W. de Bone, P. F. W. Möller, S. Baier, M. J. Degen, S. J. H. Loenen, H. P. Bartling, M. Markham, D. J. Twitchen, R. Hanson, D. Elkouss, and T. H. Taminiau, Robust quantum-network memory based on spin qubits in isotopically engineered diamond, npj Quantum Inf. **8**, 122 (2022).
[67] L. Henriet, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, Quantum computing with neutral atoms, Quantum **4**, 327 (2020).
[68] J. Ramette, J. Sinclair, Z. Vendeiro, A. Rudelis, M. Cetina, and V. Vuletić, Any-to-any connected cavity-mediated architecture for quantum computing with trapped ions or Rydberg arrays, PRX Quantum **3**, 010344 (2022).
[69] Y. Li and J. D. Thompson, High-rate and high-fidelity modular interconnects between neutral atom quantum processors, PRX Quantum **5**, 020363 (2024).
[70] C. Ryan-Anderson *et al.*, High-fidelity and fault-tolerant teleportation of a logical qubit using transversal gates and lattice surgery on a trapped-ion quantum computer, arXiv:2404.16728.
[71] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, Phys. Rev. X **11**, 041058 (2021).
[72] C. Cabrillo, J. I. Cirac, P. García-Fernández, and P. Zoller, Creation of entangled states of distant atoms by interference, Phys. Rev. A **59**, 1025 (1999).

[73] S. D. Barrett and P. Kok, Efficient high-fidelity quantum computation using matter qubits and linear optics, Phys. Rev. A **71**, 060310 (2005).
[74] V. M. Schäfer, C. J. Ballance, K. Thirumalai, L. J. Stephenson, T. G. Ballance, A. M. Steane, and D. M. Lucas, Fast quantum logic gates with trapped-ion qubits, Nature (London) **555**, 75 (2018).
[75] Y. Chew, T. Tomita, T. P. Mahesh, S. Sugawa, S. de Léséleuc, and K. Ohmori, Ultrafast energy exchange between two single Rydberg atoms on a nanosecond timescale, Nat. Photonics **16**, 724 (2022).
[76] S. Bartolucci, P. M. Birchall, M. Gimeno-Segovia, E. Johnston, K. Kieling, M. Pant, T. Rudolph, J. Smith, C. Sparrow, and M. D. Vidrighin, Creation of entangled photonic states using linear optics, arXiv:2106.13825.
[77] W. Roga, R. Ikuta, T. Horikiri, and M. Takeoka, Efficient Dicke-state distribution in a network of lossy channels, Phys. Rev. A **108**, 012612 (2023).
[78] H. Shimizu, W. Roga, D. Elkouss, and M. Takeoka, Simple loss-tolerant protocol for Greenberger-Horne-Zeilinger-state distribution in a quantum network, Phys. Rev. A **111**, 022624 (2025).
[79] F. K. Marqversen, G. Baranes, M. Sirotin, and J. Borregaard, Fault-tolerant interfaces for modular quantum computing on diverse qubit platforms, Phys. Rev. Res. **8**, 023097 (2026).
[80] C. A. Pattison, G. Baranes, J. P. B. Ataides, M. D. Lukin, and H. Zhou, Fast quantum interconnects via constant-rate entanglement distillation, arXiv:2408.15936.
[81] M. H. Abobeih, Y. Wang, J. Randall, S. J. H. Loenen, C. E. Bradley, M. Markham, D. J. Twitchen, B. M. Terhal, and T. H. Taminiau, Fault-tolerant operation of a logical qubit in a diamond quantum processor, Nature (London) **606**, 884 (2022).
[82] C. E. Bradley, J. Randall, M. H. Abobeih, R. C. Berrevoets, M. J. Degen, M. A. Bakker, M. Markham, D. J. Twitchen, and T. H. Taminiau, A ten-qubit solid-state spin register with quantum memory up to one minute, Phys. Rev. X **9**, 031045 (2019).
[83] H. K. C. Beukers, M. Pasini, H. Choi, D. Englund, R. Hanson, and J. Borregaard, Remote-entanglement protocols for stationary qubits with photonic interfaces, PRX Quantum **5**, 010202 (2024).
[84] K. Tiurev, P. L. Mirambell, M. B. Lauritzen, M. H. Appel, A. Tiranov, P. Lodahl, and A. S. Sørensen, Fidelity of time-bin-entangled multiphoton states from a quantum emitter, Phys. Rev. A **104**, 052604 (2021).
[85] P. L. Mirambell, Fidelity characterization of spin-photon entangled states (2019).
[86] H. Shibata, K. Shimizu, H. Takesue, and Y. Tokura, Ultimate low system dark-count rate for superconducting nanowire single-photon detector, Opt. Lett. **40**, 3428 (2015).
[87] N. Delfosse and N. H. Nickerson, Almost-linear time decoding algorithm for topological codes, Quantum **5**, 595 (2021).
[88] www.surf.nl.
[89] S. Singh, R. Kashiwagi, K. Tanji, W. Roga, D. Bhatti, M. Takeoka, and D. Elkouss, Source code and data underlying the publication "fault-tolerant modular quantum computing with surface codes using single-shot emission-based hardware" (2025).
[90] H. Cao, S. Zhao, D. Feng, Z. Shen, H. Yan, T. Su, W. Sun, H. Xu, F. Pan, H. Yu, and P. Zhang, Exact decoding of repetition code under circuit level noise, arXiv:2501.03582.
[91] O. Higgott, T. C. Bohdanowicz, A. Kubica, S. T. Flammia, and E. T. Campbell, Improved decoding of circuit noise and fragile boundaries of tailored surface codes, Phys. Rev. X **13**, 031007 (2023).
[92] D. J. Reilly, Engineering the quantum-classical interface of solid-state qubits, npj Quantum Inf. **1**, 15011 (2015).
[93] A. Gold, J. P. Paquette, A. Stockklauser, M. J. Reagor, M. S. Alam, A. Bestwick, N. Didier, A. Nersisyan, F. Oruc, A. Razavi, B. Scharmann, E. A. Sete, B. Sur, D. Venturelli, C. J. Winkleblack, F. Wudarski, M. Harburn, and C. Rigetti, Entanglement across separate silicon dies in a modular superconducting qubit device, npj Quantum Inf. **7**, 142 (2021).
[94] O. Crawford, J. R. Cruise, N. Mertig, and M. F. Gonzalez-Zalba, Compilation and scaling strategies for a silicon quantum processor with sparse two-dimensional connectivity, npj Quantum Inf. **9**, 13 (2023).
[95] M. W. Doherty, N. B. Manson, P. Delaney, F. Jelezko, J. Wrachtrup, and L. C. L. Hollenberg, The nitrogen-vacancy colour centre in diamond, Phys. Rep. **528**, 1 (2013).
[96] E. Togan, Y. Chu, A. S. Trifonov, L. Jiang, J. Maze, L. Childress, M. V. G. Dutt, A. S. Sørensen, P. R. Hemmer, A. S. Zibrov, and M. D. Lukin, Quantum entanglement between an optical photon and a solid-state spin qubit, Nature (London) **466**, 730 (2010).
[97] H. Bernien, B. Hensen, W. Pfaff, G. Koolstra, M. S. Blok, L. Robledo, T. H. Taminiau, M. Markham, D. J. Twitchen, L. Childress, and R. Hanson, Heralded entanglement between solid-state qubits separated by three metres, Nature (London) **497**, 86 (2013).
[98] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, High-fidelity projective read-out of a solid-state spin quantum register, Nature (London) **477**, 574 (2011).
[99] B. Hensen, H. Bernien, A. E. Dréau, A. A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, Nature (London) **526**, 682 (2015).
[100] N. Kalb, A. A. Reiserer, P. C. Humphreys, J. J. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson, Entanglement distillation between solid-state quantum network nodes, Science **356**, 928 (2017).
[101] P. C. Humphreys, N. Kalb, J. P. J. Morits, R. N. Schouten, R. F. L. Vermeulen, D. J. Twitchen, M. Markham, and R. Hanson, Deterministic delivery of remote entanglement on a quantum network, Nature (London) **558**, 268 (2018).
[102] F. Rozpędek, R. Yehia, K. Goodenough, M. Ruf, P. C. Humphreys, R. Hanson, S. Wehner, and D. Elkouss, Near-term quantum-repeater experiments with nitrogen-vacancy

centers: Overcoming the limitations of direct transmission, Phys. Rev. A **99**, 052330 (2019).

[103] S. L. N. Hermans, M. Pompili, L. D. Santos Martins, A. R-P Montblanch, H. K. C. Beukers, S. Baier, J. Borregaard, and R. Hanson, Entangling remote qubits using the single-photon protocol: An in-depth theoretical and experimental study, New J. Phys. **25**, 013011 (2023).

[104] K. Nemoto, M. Trupke, S. J. Devitt, A. M. Stephens, B. Scharfenberger, K. Buczak, T. Nöbauer, M. S. Everitt, J. Schmiedmayer, and W. J. Munro, Photonic architecture for scalable quantum information processing in diamond, Phys. Rev. X **4**, 031022 (2014).

[105] L. Robledo, H. Bernien, Tvan der Sar, and R. Hanson, Spin dynamics in the optical cycle of single nitrogen-vacancy centres in diamond, New J. Phys. **13**, 025013 (2011).

[106] C. K. Hong, Z. Y. Ou, and L. Mandel, Measurement of subpicosecond time intervals between two photons by interference, Phys. Rev. Lett. **59**, 2044 (1987).

[107] A. Kiraz, M. Atatüre, and A. Imamoğlu, Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing, Phys. Rev. A **69**, 032305 (2004).

[108] A. Dahlberg, M. Skrzypczyk, T. Coopmans, L. Wubben, F. Rozpundefineddek, M. Pompili, A. Stolk, P. Pawełczak, R. Knegjens, J. de Oliveira Filho, R. Hanson, and S. Wehner, A link layer protocol for quantum networks, in *Proceedings of the ACM Special Interest Group on Data Communication, SIGCOMM '19* (Association for Computing Machinery, New York, NY, USA, 2019), pp. 159–173.

[109] D. Riedel, I. Söllner, B. Shields, S. Starosielec, P. Appel, E. Neu, P. Maletinsky, R. J. Warburton, M. Agio, K. Srinivasan, and C. Soci, Deterministic enhancement of coherent photon generation from a nitrogen-vacancy center in ultrapure diamond, Phys. Rev. X **7**, 031040 (2017).

[110] I. Aharonovich, A. D. Greentree, and S. Prawer, Diamond photonics, Nat. Photonics **5**, 397 (2011).

[111] R. Albrecht, A. Bommer, C. Deutsch, J. Reichel, and C. Becher, Coupling of a single nitrogen-vacancy center in diamond to a fiber-based microcavity, Phys. Rev. Lett. **110**, 243602 (2013).

[112] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).

[113] A. Faraon, C. Santori, Z. Huang, V. M. Acosta, and R. G. Beausoleil, Coupling of nitrogen-vacancy centers to photonic crystal cavities in monocrystalline diamond, Phys. Rev. Lett. **109**, 033604 (2012).

[114] B. J. M. Hausmann, B. Shields, Q. Quan, P. Maletinsky, M. McCutcheon, J. T. Choy, T. M. Babinec, A. Kubanek, A. Yacoby, M. D. Lukin, and M. Lončar, Integrated diamond networks for quantum nanophotonics, Nano Lett. **12**, 1578 (2012).

[115] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, Detecting single infrared photons with 93% system efficiency, Nat. Photonics **7**, 210 (2013).

[116] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, Invited review article: Single-photon sources and detectors, Rev. Sci. Instrum. **82**, 071101 (2011).

[117] J. Sperling, W. Vogel, and G. S. Agarwal, True photocounting statistics of multiple on-off detectors, Phys. Rev. A **85**, 023820 (2012).

[118] A. E. Lita, A. J. Miller, and S. W. Nam, Counting near-infrared single-photons with 95% efficiency, Opt. Express **16**, 3032 (2008).

[119] F. Mattioli, Z. Zhou, A. Gaggero, R. Gaudio, S. Jahanmirinejad, D. Sahin, F. Marsili, R. Leoni, and A. Fiore, Photon-number-resolving superconducting nanowire detectors, Supercond. Sci. Technol. **28**, 104001 (2015).

[120] T. Gerrits, N. Thomas-Peter, J. C. Gates, A. E. Lita, B. J. Metcalf, B. Calkins, N. A. Tomlin, A. E. Fox, A. L. Linares, J. B. Spring, N. K. Langford, R. P. Mirin, P. G. R. Smith, I. A. Walmsley, and S. W. Nam, On-chip, photon-number-resolving, telecommunication-band detectors for scalable photonic information processing, Phys. Rev. A **84**, 060301 (2011).

[121] G. Homa, A. Ortega, and M. Koniorczyk, Choi representation of completely positive maps in brief, Z. Naturforschung A **79**, 1123 (2024).

[122] M. Frembs and E. G. Cavalcanti, Variations on the Choi—Jamiołkowski isomorphism, J. Phys. A **57**, 265301 (2024).

[123] A. Katabarwa, A dynamical interpretation of the Pauli twirling approximation and quantum error correction, arXiv:1701.03708.

[124] M. R. Geller and Z. Zhou, Efficient error models for fault-tolerant architectures and the Pauli twirling approximation, Phys. Rev. A **88**, 012314 (2013).

[125] Y. Wu, N. Liyanage, and L. Zhong, An interpretation of union-find decoder on weighted graphs, arXiv:2211.03288.

[126] T. Chan and S. C. Benjamin, Actis: A strictly local union—Find decoder, Quantum **7**, 1183 (2023).

[127] J. Edmonds, Paths, trees, and flowers, Can. J. Math. **17**, 449 (1965).

[128] O. Higgott and C. Gidney, Sparse blossom: Correcting a million errors per core second with minimum-weight matching, Quantum **9**, 1600 (2025).

[129] C. Wang, J. Harrington, and J. Preskill, Confinement-Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory, Ann. Phys. **303**, 31 (2003).

[130] A. Ruckstuhl, *Introduction to Nonlinear Regression* (IDP Institut für Datenanalyse und Prozessdesign, ZHAW Zürcher Hochschule für Angewandte Wissenschaften, 2010).

[131] D. Constales, G. S. Yablonsky, D. R. D'hooge, J. W. Thybaut, and G. B. Marin, Chapter 9—Experimental data analysis: Data processing and regression, in *Advanced Data Analysis & Modelling in Chemical Engineering* (Elsevier, Amsterdam, 2017), pp. 285–306.